

2021/2022

Spécialité : ERP-BI

**Développement d'une solution
décisionnelle et de prévision d'un SGBD
Oracle**

- Réalisé par : **Zaki Bouzidi**
- Encadré par :
 - Encadrant ESPRIT: Amine Zribi
 - Encadrant Entreprise: Wissem Brahmi

Encadrant Professionnel



Signature et cachet

Encadrant Académique



Signature et cachet

Dédicace

“

*Toutes les lettres ne sauraient trouver les mots qu'il faut. . .
Tous les mots ne sauraient exprimer la gratitude, l'amour, Le
respect, la reconnaissance. . .*

À mes chers parents Ammar et Maryem

*Aucune dédicace ne saurait exprimer mon respect, mon amour
éternel et ma considération pour les sacrifices que vous avez
consenti pour mon instruction et mon bien être. Je vous
remercie pour tout le soutien et l'amour que vous me portez
depuis mon enfance et j'espère que votre bénédiction
m'accompagne toujours. Puisse Dieu vous accorder santé,
bonheur et longue vie et faire en sorte que jamais je ne vous
déçoive Vous avez toujours été mon école de patience, de
confiance et surtout d'espoir et d'amour.*

À mes deux sœurs et mon petit frère

*Qui je le sais ma réussite est très importante pour vous. Que
Dieu vous paye pour tous vos bienfaits, pour son sacrifice, son
amour et son encouragement qui m'ont guidé vers la réussite.*

”

Zaki Bouzidi

Remerciements

En premier lieu, je tiens à remercier Dieu le tout puissant pour Sa faveur et Sa grâce.

Ensuite, à tous mes professeurs et cadre enseignant qui ont guidé mon tout premier pas à l'université. Leurs conseils, leurs leçons et leurs critiques m'ont été d'un grand apport tant dans la transmission de savoir que dans le développement de mes compétences.

J'adresse une sincère reconnaissance et une immense gratitude à tout le personnel de la société WorldSoft Group et plus particulièrement à M. Mohamed Walid Medini (IT Manager) et M. Wissem Brahmi pour leur immense qualité professionnelle, mais aussi humaine qui m'ont aidé de venir à bout de ce modeste travail dans la confiance et la sérénité.

J'ai l'insigne honneur de remercier mon encadrant universitaire M. Amine Zribi qui trace une fresque vivante des grandes enseignantes qui ont bien servi et servent encore le savoir. Son dévouement, son labeur incessant et ce ne sont pas là minces mérites lui ont procuré le respect de tous.

A mes chers professeurs membres du jury qui m'honorent de leur élogieuse présence veuillez bien trouver ici l'expression de ma reconnaissance et de mon indéfectible respect.

A ce titre, l'hommage que je tiens à rendre à vous tous n'est que le faible témoignage de la grande dette dont je vous suis redevable.

Table des matières

INTRODUCTION GENERALE	1
Chapitre 1: Etude Préliminaire	3
Introduction	3
1. Présentation de l'organisme d'accueil	3
1.1 Description de la société	3
1.2 Organigramme de la société	3
2. Présentation du projet	4
3. Etude préalable	4
3.1 Etude de l'existant	4
3.2 Problématique	5
3.3 La solution proposée	5
4. Choix de la méthodologie du travail	6
4.1 La méthodologie Agile	6
4.2 Les principales méthodes Agiles	8
4.3 Description de la méthode agile Scrum	8
Conclusion	10
Chapitre 2: Etat de l'art de l'informatique décisionnelle	11
Introduction	11
1. Le système décisionnel	11
1.1 Le système d'information dans l'entreprise	11
1.2 La Business Intelligence (Informatique décisionnelle)	12
2. Modélisation	12
2.1 Base de données	13
2.2 L'entrepôt de données	13
2.3 Le processus ETL	13
2.4 Etude comparative des outils ETL	15
2.5 Modélisation multidimensionnelle	17
2.6 Analyse	20
2.6.1 On-Line Analytical Processing (OLAP)	20
2.7 La visualisation	21
2.8 Déploiement	22
Conclusion	23
Chapitre 3 : Sprint 0 - Analyse et spécification des besoins	24

Introduction	24
1. Spécification des besoins	24
1.1 Identification d'acteurs	24
1.2 Analyse des besoins fonctionnels	24
2. Pilotage du projet par Scrum	26
2.1 Identification de l'équipe Scrum	26
2.2 Elaboration de Product Backlog	26
2.3 Découpage du projet.....	28
Conclusion.....	28
Chapitre 4: Sprint 1 – Conception du Data Warehouse et Développement du ETL	29
Introduction	29
1. Conception du Data Warehouse	29
1.1 Choix des dimensions.....	29
1.2 Le choix des mesures	32
1.3 Table de fait « Fact_Sales »	32
1.4 Structuration du Data Warehouse.....	33
2. Conception de l'ETL (Extract – Transform – Load)	34
2.1 Etude de données.....	35
2.2 Réalisation de l'ETL	35
Chapitre 5: Sprint 2 – Réalisation du Tableau de bord et Phase de clôture	40
Introduction	40
1. Environnement de travail.....	40
1.1 Environnement matériel	40
2. Environnement logiciel et langage de développement	40
2.1 Environnement logiciel	40
2.2 Langage de développement	42
3. Réalisation du tableau de bord	42
3.1 Connexion	42
3.2 Visualisation du tableau de bord	44
3.2 Déploiement sur Web.....	49
Conclusion	51
CONCLUSION GENERALE.....	52
Webographie	53

Liste des figures

Figure 1: Logo de la Société WorldSoft.....	3
Figure 2: Organigramme de la société WorldSoft.....	4
Figure 3: Dashboard de l'existant (SRA).....	5
Figure 4: Les valeurs de la méthode Agile.....	6
Figure 5: Les principes de la méthode agile.....	7
Figure 6: Le processus Scrum	9
Figure 7: Backlog du Produit	10
Figure 8: Sprint Backlog	10
Figure 9: Processus ETL (Extract Transform Load).....	14
Figure 10: Logo Talend et Informatica	15
Figure 11: Schéma en étoile	17
Figure 12: Schéma en flocon de neige	18
Figure 13: Schéma en constellation.....	18
Figure 14: Illustration d'une table de Faits	19
Figure 15: Différence entre Slice et Dice	21
Figure 16: Présentation de l'équipe Scrum	26
Figure 17: Diagramme de cas d'utilisation	27
Figure 18: Découpage du projet	28
Figure 19: Modélisation de l'entrepôt de données.....	34
Figure 20: Interface de connexion à la base de données	35
Figure 21: Interface 1 de la connexion de Talend	36
Figure 22: Connexion entre Talend et SQL Developer.....	37
Figure 23: Récupération des tables.....	37
Figure 24: Création d'u Job	38
Figure 25: Job crée et opérationnel	38
Figure 26: Remplissage des tables de dimension	39
Figure 27: Interface 1 connexion Power BI à Oracle	43
Figure 28: Interface 2 de connexion Power BI à Oracle	43
Figure 29: Récupération des tables sur Power BI	44
Figure 30: Ecran d'accueil de l'application.....	45
Figure 31: Rubrique vue d'ensemble	46
Figure 32: Rubrique des Agences.....	47
Figure 33: Rubrique Airlines.....	48
Figure 34: Rubrique des GDS	48
Figure 35: Rubrique des Pays.....	49
Figure 36: Authentification	50
Figure 37: Application BI sur Web	50

Liste des tableaux

Tableau 1: Etude Comparative de Talend vs Informatica	16
Tableau 2: Tableau du Product backlog	27
Tableau 3: Table de dimension Airlines.....	30
Tableau 4: Table de dimension Client.....	30
Tableau 5: Table de dimension Pays (Country)	31
Tableau 6: Table de dimension entité.....	31
Tableau 7: Table de dimension Catégorie client	31
Tableau 8: Table de dimension Global Distribution System (GDS).....	32
Tableau 9: Table des Mesures	32
Tableau 10: Description de la table de Fait	33
Tableau 11: Environnement logiciel	41
Tableau 12: Langage de développement	42

Liste des Acronymes

BI : Business Intelligence

DW : Data Warehouse

ED : Entrepôt de Données

ETL : Extract Transform Load (Extraction, Transformation et Chargement)

GDS : Global Distribution System (Système global de distribution)

IATA : International Air Transport Association (Association Internationale de Transport Aérien)

SGBD : Système de Gestion de Base de Données

SQL : Structured Query Language (Language de requête structurée)

INTRODUCTION GENERALE

Avec une concurrence croissante et une évolution constante du monde des affaires, les organisations sont de plus en plus soucieuses de contrôler leurs activités financières, commerciales et industrielles, tout en gardant un œil sur leur gestion d'informations. Les entreprises cherchent à surveiller de près le marché pour ne pas se laisser distancées par leurs concurrents et répondre rapidement aux attentes de leurs clients et partenaires. Les managers doivent être en mesure d'accomplir les tâches qui leur sont assignées, quel que soit leur domaine d'activité. Ils doivent prendre des décisions et des initiatives opportunes qui influenceront en partie la stratégie et l'avenir de l'entreprise. Ces décisions résultent en des conséquences importantes pour la survie de l'entreprise et doivent être prises de manière réfléchie, sans être trop prématurées.

Aujourd'hui, la gestion des systèmes d'information est devenue un processus incontournable pour toutes les entreprises. Cependant, il est difficile de savoir comment identifier et présenter les informations à ceux qui en ont besoin, surtout lorsque les entreprises sont submergées par une quantité énorme de données provenant de différentes sources. Ce projet vise à mettre en place un système capable de consolider les données provenant des systèmes traditionnels, tels que les bases de données relationnelles, afin de fournir des informations de qualité aux décideurs. L'objectif est de leur mettre à la disposition des informations éclairantes pour faciliter une prise de décision prompte en connaissance de cause.

La mise en place d'un entrepôt de données fiable est donc indispensable pour un tel système décisionnel. Dans ce contexte, la société "WorldSoft Group" a proposé un projet dans le cadre d'un stage de fin d'étude en vue de l'obtention d'un diplôme d'ingénierie en informatique. Le projet consiste à concevoir et développer un outil décisionnel en utilisant des outils d'extraction, de transformation et de chargement (ETL) à partir des informations de la société.

Ce rapport est composé de cinq parties. Le premier chapitre, intitulé "Etude préliminaire", présente la société dans laquelle le stage a été effectué, suivi d'une brève présentation du projet. Ensuite, une étude critique de l'existant est effectuée pour proposer une solution, puis une méthode est définie pour guider le rapport tout au long du processus.

Le deuxième chapitre, "Etat de l'art de l'informatique décisionnelle", définit les concepts clés de la *Business Intelligence* et de l'entrepôt de données.

Le troisième chapitre, "Analyse et Spécification des besoins", commence par identifier les besoins fonctionnels et non fonctionnels, suivi de la présentation de l'équipe de pilotage du projet et de la découpe du projet.

Le quatrième chapitre, "Conception du *Data Warehouse* et développement de l'ETL", traite la conception de l'entrepôt de données et le développement de l'ETL.

Enfin, le cinquième chapitre traitera "Le déploiement et la réalisation du tableau de bord".

Chapitre 1: Etude Préliminaire

Introduction

Dans ce chapitre, je commence par la présentation rapide de l'entreprise dans laquelle j'ai effectué mon stage, puis j'introduirai le sujet de thèse. Enfin, je mène une étude critique de l'état actuel des choses.

1. Présentation de l'organisme d'accueil

Pour situer correctement le travail dans son contexte, il est important de présenter l'entreprise qui m'a accueillie pour mon stage de fin d'étude « WorldSoft Group».



Figure 1: Logo de la Société WorldSoft

1.1 Description de la société

WorldSoft Group est une entreprise tunisienne de développement informatique fondée en 2014. Son domaine d'expertise se concentre sur la conception des plateformes web et mobiles de réservation en ligne.

Son objectif principal est de proposer un espace de vente de produits touristiques en ligne, incluant des modules de réservation pour les vols, les hôtels, les transferts,... La société est implantée géographiquement dans deux pays : en Tunisie, à Tunis plus précisément dans le quartier de Marsa 2046, Sidi Daoud Marsa Mall, et en Algérie, à Alger.

1.2 Organigramme de la société

Le schéma organisationnel de WorldSoft Group illustre les différentes fonctions au sein de l'entreprise. Il est constitué d'une direction générale et de trois principaux départements, à savoir le département financier, le département de projet qui comprend une équipe de développement, et enfin

le département des ressources humaines. Ce schéma permet de mieux comprendre la structure de la société et les relations hiérarchiques entre les différents services.

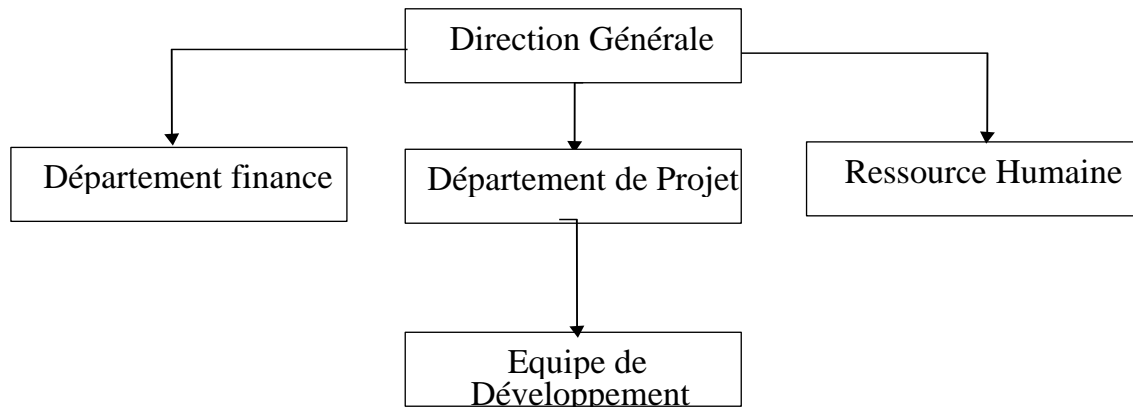


Figure 2: Organigramme de la société WorldSoft

1. Présentation du projet

Ce projet a pour finalité de permettre à la société « WorldSoft Group » d'étendre son champ d'action en introduisant un module de BI dans ses services. L'objectif de ce module est de permettre aux responsables de suivre les résultats de leurs activités et de faciliter les prises de décision relatives à l'orientation de l'activité de la société. Dans ce contexte, la mission était de développer un système décisionnel et de prévision qui permettra de contrôler l'évolution du chiffre d'affaires de la société. Ce projet s'inscrit dans le cadre de notre formation d'ingénieurs en informatique à l'ESPRIT, et a été réalisé au sein des locaux de la société « WorldSoft Group ».

2. Etude préalable

L'analyse préliminaire de notre projet commence par une étude approfondie de l'existant. Cette étape est cruciale car elle nous permet de comprendre en détail les besoins actuels de la société et d'identifier les problèmes et les défis auxquels notre système doit faire face. En examinant les processus métier existants, nous pouvons déterminer les points faibles du système actuel, les éventuels problèmes de performance et les goulots d'étranglement. Cette analyse nous fournira une base solide pour développer un système BI performant et efficace pour « WorldSoft Groupe ».

3.1 Etude de l'existant

La société « WorldSoft » m'a fourni un système existant nommé « SRA » que j'ai analysé. Cette solution présente certains avantages tels que son ergonomie, mais elle présente également des lacunes

en termes de sécurité et de génération de données dynamiques, ainsi que la présentation des activités par zone ou pays. La figure 3 illustre le tableau de bord de l'existant.

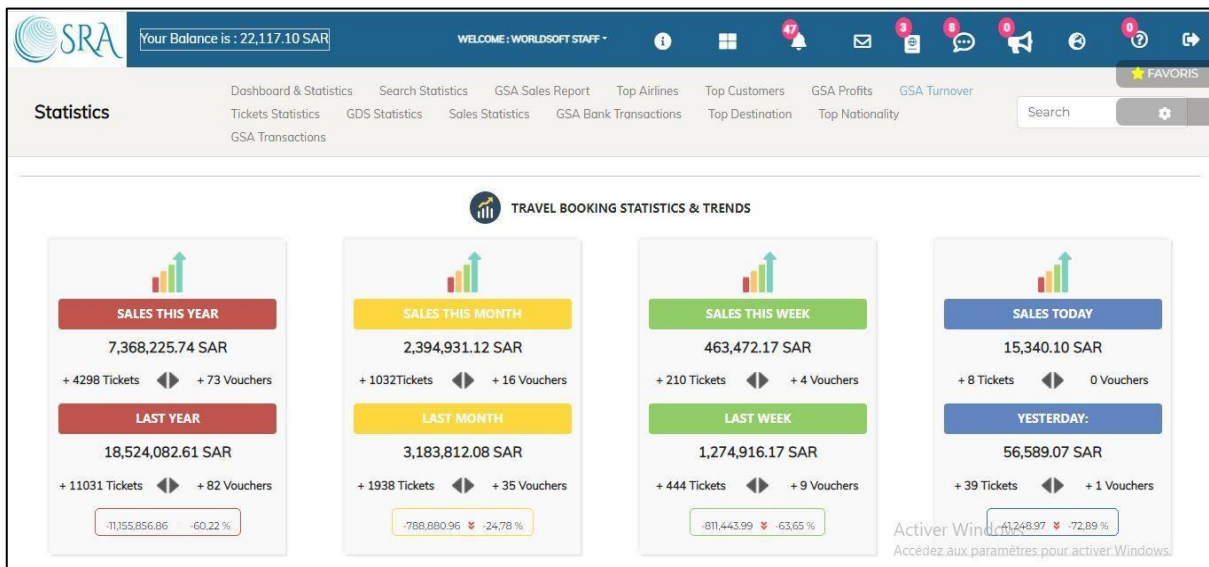


Figure 3: Dashboard de l'existant (SRA)

3.2 Problématique

Dans ce contexte, mon projet vise à mettre en place un système décisionnel et de prévision basé sur la BI pour la société « WorldSoft ». Ce système permettra de collecter, stocker, traiter et communiquer les informations de manière plus efficace et performante, tout en réduisant les temps de réponse et en améliorant la sécurité des données. Grâce à cette solution, les décideurs auront accès à des informations pertinentes, fiables et en temps réel, ce qui facilitera leur prise de décision et orientera l'activité de l'entreprise.

Notre système offrira également la possibilité de générer des rapports automatisés pour une vue globale de l'activité de la société. En somme, notre projet vise à améliorer la gestion des informations au sein de la société « WorldSoft » et à contribuer à sa croissance.

3.3 La solution proposée

Ces objectifs sont tout à fait pertinents pour répondre aux besoins de la société WorldSoft. Le Data Warehouse permettra de collecter et de stocker les données de manière structurée et centralisée, facilitant ainsi leur analyse et leur exploitation. Le cube OLAP, quant à lui, permettra d'effectuer des analyses multidimensionnelles, offrant ainsi une vision plus complète et plus approfondie des données. La connexion à Power BI permettra de créer des tableaux de bord interactifs et dynamiques, facilitant ainsi la prise de décision. Cette solution répondra donc aux besoins des décideurs en leur fournissant des informations claires et précises pour une meilleure orientation des activités de la société.

3. Choix de la méthodologie du travail

Tout à fait, le choix d'une méthodologie de gestion de projet est essentiel pour garantir la réussite et l'efficacité du projet. Il existe plusieurs méthodologies telles que Waterfall, Agile, Scrum, Kanban,... Le choix dépend généralement des spécificités du projet, de ses objectifs, des ressources disponibles et des contraintes.

Il est important de choisir une méthode adaptée à notre projet afin de nous aider à organiser et planifier les différentes étapes du projet, ainsi qu'à gérer les ressources humaines, les coûts et les délais.

Il est également important de noter que la méthode choisie doit être flexible et permet d'adapter le projet en fonction des imprévus et des changements, tout en gardant le cap sur les objectifs à atteindre.

4.1 La méthodologie Agile

Une méthode agile est une approche collaborative, itérative et incrémentale, qui se caractérise par une faible dose de formalisme. Elle permet de produire des résultats de qualité en tenant compte de l'évolution des besoins des clients. Elle repose sur des cycles de travail courts et permet de découper le projet en blocs plus petits et hiérarchisés en fonction des priorités. Les méthodes agiles se fondent sur des valeurs (figure 4) et des principes (figure 5).



Figure 4: Les valeurs de la méthode Agile

Les quatre valeurs des méthodes agiles sont :

- L'individu et les interactions plus que les processus et les outils : cela signifie que la communication et la collaboration entre les membres de l'équipe sont plus importantes que les processus et les outils utilisés.

- Le logiciel fonctionnel plus qu'une documentation exhaustive : cela signifie que le produit final est plus important que la documentation exhaustive qui l'accompagne.
- La collaboration avec le client plus que la négociation contractuelle : cela signifie que la collaboration avec le client est primordiale pour atteindre les objectifs du projet.
- La réponse au changement plus que le suivi d'un plan : cela signifie que la capacité de s'adapter aux changements est cruciale pour réussir dans un environnement en constante évolution.

Ainsi, je cite au niveau de la figure 5 les principes de la méthodologie Agile.

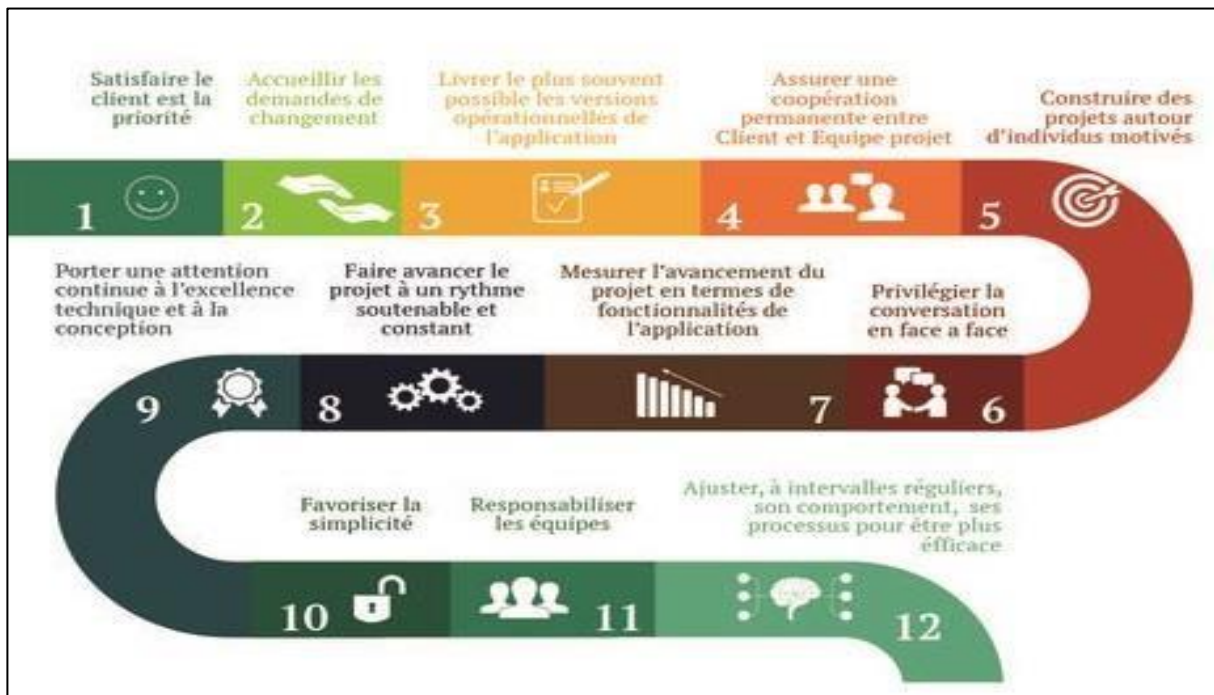


Figure 5: Les principes de la méthode agile

Nous distinguons alors douze principes de la méthode agile qui sont :

- La satisfaction du client prime sur tout le reste
- Livrer rapidement des logiciels fonctionnels
- Accueillir les changements de besoins
- Travailler en collaboration avec le client
- Bâtir des projets autour de personnes motivées
- Privilégier la communication face-à-face
- Privilégier les logiciels fonctionnels à la documentation
- Progresser à un rythme soutenable
- Accorder de l'importance à l'excellence technique

- Simplifier le processus de développement
- Favoriser l'auto-organisation de l'équipe
- Réfléchir régulièrement à l'amélioration.

4.2 Les principales méthodes Agiles

Une fois que nous décidons d'adopter une gestion de développement Agile, il reste encore à choisir la méthodologie la plus adaptée à son projet. En effet, les méthodes Agiles sont nombreuses et peuvent être source de confusion.

De nos jours, les méthodes Agiles les plus populaires en usage sont :

- L'Extrême Programming (XP)
- Scrum
- Feature Driven Development
- Agile Unified Process (Agile UP ou AP)
- Crystal
- Dynamic System Development Method (DSDM).

Dans notre cas, nous avons décidé d'adopter la **méthode agile Scrum**, parce qu'elle est la méthode adaptée pour la réalisation du projet en équipe dont le client est impliqué dans la réalisation.

4.3 Description de la méthode agile Scrum

Le Scrum permet à une équipe de collaborer efficacement dans le but de développer un produit de meilleure qualité tout en respectant les délais. La construction du produit final se fait en plusieurs itérations nommées « Sprint » d'une durée de 2 à 4 semaines. Durant chacune de ces itérations, une partie du produit nommée « Incrément » est réalisée se basant sur les parties créées lors des itérations précédentes et livrées à la fin du Sprint.

A chaque fin de Sprint, il est possible de voir fonctionner le produit courant et décider soit de le livrer, soit de le continuer et améliorer pendant un autre sprint.

Le processus de Scrum se présente comme suit dans la figure 6.



Figure 6: Le processus Scrum

Le Scrum définit trois principaux rôles à savoir :

- **Product Owner** : qui porte la vision du produit à réaliser et travaille en interaction avec l'équipe du développement. Il s'agit généralement d'un expert du domaine métier du projet.
- **Le Scrum Master** : qui doit maîtriser le Scrum et s'assurer que ce dernier est correctement appliqué. Pour cela, il joue le rôle de coach à la fois auprès du Product Owner et auprès de l'équipe de développement.
- **L'équipe de développement** : est chargée de transformer les besoins exprimés par le Product Owner en fonctionnalités utilisables. Elle est pluridisciplinaire et peut donc encapsuler d'autres rôles tels que développeur, graphiste, ingénieur système,...

4.3.1 Les artefacts de Scrum

Les artefacts sont des éléments permettant de faire fonctionner le cadre de travail Scrum, en accord avec les principes et les pratiques s'inscrivant dans processus de l'agilité.

Les artefacts sont regroupés au nombre de trois:

- **Le Backlog de Produit (Product Backlog)** qui est une liste ordonnée et priorisée de toutes les fonctionnalités qui doivent constituer le produit. Il contient tous les besoins recueillis pour faire évoluer le produit. Comme les besoins du client ne sont pas stables, cette liste évolue sans cesse avec les changements faits au cours du projet. Ce dynamisme présente les avantages majeurs du Scrum.



Figure 7: Backlog du Produit

- **Sprint Backlog (Carnet de Sprint)** est une liste contenant les spécifications techniques pour les tâches devant être accomplies par l'équipe de développement au cours du Sprint bien déterminé. Il est mis à jour régulièrement par l'équipe dans le but d'avoir une vision claire et précise sur l'avancement du Sprint.



Figure 8: Sprint Backlog

- **Burn Down Chart graphique d'avancement** est une représentation graphique qui permet à une équipe travaillant sur un projet bien déterminé de visualiser la progression de leur travail par rapport à ce qui était planifiée au début du Sprint. Le Burn Down Chart est mis à jour régulièrement par le Scrum dans le but de prendre de mesure nécessaire.

Conclusion

Le premier chapitre de mon rapport a débuté par une introduction générale du projet, comprenant une présentation de l'entreprise d'accueil. J'ai ensuite identifié la problématique en étudiant les outils existants sur le marché, proposant ainsi une solution appropriée. Et également j'ai fait le choix de la méthodologie à suivre tout au long du projet.

Le chapitre suivant se concentrera sur l'analyse des besoins.

Chapitre 2: Etat de l'art de l'informatique décisionnelle

Introduction

Le deuxième chapitre du travail porte sur l'informatique décisionnelle en entreprise, en abordant les différentes étapes telles que le stockage, la modélisation, l'analyse et la visualisation des données. Nous procédons également à une comparaison des outils ETL.

L'objectif de ce chapitre est de clarifier certaines notions clés liées aux systèmes décisionnels.

1. Le système décisionnel

1.1 Le système d'information dans l'entreprise

Le système d'information est un ensemble organisé de ressources qui a pour rôle de collecter, stocker, traiter et distribuer des informations. Dans le contexte d'une entreprise, le système d'information a pour objectif de faciliter la mise en œuvre de la stratégie en soutenant les activités de l'entreprise. Il est conçu en fonction des exigences métiers et des processus définis par l'entreprise, et est composé de moyens humains, logiciels et matériels.

Il est d'usage de distinguer trois types différents de système d'information dont :

- **Le système opérant** est le fondement de toute organisation, car il permet la transformation de l'information pour la restituer aux personnes concernées. Il englobe les différents services de l'entreprise qui sont chargés de réaliser les activités opérationnelles nécessaires pour atteindre les objectifs fixés.
- **Le système de pilotage** est chargé de diriger et de contrôler le système opérant. Il occupe donc une position hiérarchique supérieure dans le système d'information, en fixant les objectifs et en prenant les décisions stratégiques pour l'entreprise.
- **Le système d'information** joue un rôle central entre les systèmes opérant et de pilotage d'une entreprise. Son rôle est de collecter, stocker, traiter et distribuer des données et des informations utiles aux deux autres systèmes. Il assure ainsi la communication et la coordination entre les différents services de l'entreprise, permettant une prise de décision efficace et une meilleure gestion des activités.

Le système d'information classique est un mode de fonctionnement des entreprises qui présente plusieurs inconvénients. Tout d'abord, il souffre souvent d'un manque de culture de partage de l'information, ce qui limite la circulation de la connaissance au sein de l'entreprise. De plus, la gestion de l'information peut être négligée, ce qui peut entraîner un manque de temps pour codifier les informations et les connaissances, et un refus de l'apprentissage et du désapprentissage. Ces facteurs peuvent entraver le processus de gestion globale de l'entreprise.

1.2 La Business Intelligence (Informatique décisionnelle)

La *Business Intelligence* se réfère à un processus technologique qui permet d'analyser les données et de présenter les informations pour aider les dirigeants, managers et utilisateurs finaux à prendre des décisions éclairées pour l'entreprise.

Elle comprend une variété d'outils, d'applications et de méthodologies qui permettent aux organisations de collecter des données à partir de sources internes et externes, de les préparer pour l'analyse et de créer des rapports, tableaux de bord et autres outils de visualisation de données pour rendre les résultats analytiques disponibles aux décideurs et opérations.

Les entreprises utilisent de plus en plus les logiciels de *Business Intelligence* pour extraire des informations précieuses à partir de grandes quantités de données qu'elles stockent, afin de détecter les opportunités de business, d'optimiser les campagnes marketing ou financières. Au fil des années, la *Business Intelligence* est devenue une fonction clé dans toutes les entreprises, passant d'un domaine principalement lié à l'IT à une fonction transversale et stratégique.

2. Modélisation

Actuellement, les outils de *Business Intelligence* sont devenus plus flexibles, grâce aux différents matériels, logiciels, appareils et services disponibles pour collecter des données structurées et non structurées provenant de sources variées dans différents formats.

Les outils BI modernes sont capables de traiter plusieurs types d'analyse à savoir :

- **La technique de l'analyse de contenu** est une application de la BI qui permet d'analyser des données non structurées telles que les commentaires sur les médias sociaux ou les courriels. En combinant ces informations avec des données structurées, les entreprises peuvent recueillir des idées utiles pour leur activité.
- **La technique de l'analyse d'entreprise** en examinant les modèles d'affaires passés, présents et futurs pour déterminer ce qui fonctionne bien et ce qui peut être amélioré pour augmenter l'efficacité

et la rentabilité de l'entreprise. Cette analyse peut être utilisée pour identifier les tendances et les opportunités, et pour prendre des décisions stratégiques éclairées en vue d'une croissance et d'une réussite à long terme.

2.1 Base de données

Une base de données peut être définie comme un ensemble structuré de données organisées et stockées de manière à être facilement accessibles, gérées et mises à jour. Les bases de données sont généralement associées à un SGBD, qui permet de gérer et de manipuler les données stockées dans la base de données. Les SGBD sont des logiciels informatiques tels que SQL Server, Oracle SQL Developer, Oracle PL/SQL Developer et MySQL, qui permettent de stocker, de consulter, de mettre à jour et de structurer les données, ainsi que de les partager avec d'autres applications informatiques. Les SGBD sont couramment utilisés dans une variété d'applications informatiques, telles que les logiciels de réservation, les guichets automatiques bancaires,...

2.2 L'entrepôt de données

L'entrepôt de données est une base de données spécialement conçue pour collecter, organiser, enregistrer et stocker des informations provenant de diverses sources de données opérationnelles, dans le but de fournir une référence unique pour aider à la prise de décision en entreprise. Les données sont souvent utilisées pour générer des statistiques et des rapports grâce à des outils de *reporting*. De nos jours, de nombreuses entreprises modernes utilisent des entrepôts de données pour faciliter la prise de décisions et atteindre leurs objectifs.

Les tâches pouvant être effectuées par un entrepôt de données sont :

- Suivre, gérer et améliorer la performance de l'entreprise
- Optimiser la logistique et les opérations
- Améliorer l'efficacité en matière de gestion de produits et de développement
- Améliorer le relation-client
- Nettoyer et améliorer la qualité des données de l'entreprise.

2.3 Le processus ETL

Un ETL (*Extract Transform and Load*) est un processus visant à extraire des données de diverses sources, les transformer pour les rendre compatibles entre elles, puis les charger dans un entrepôt de données (*Data Warehouse*). Les entreprises utilisent l'ETL pour faciliter la manipulation, le traitement et l'analyse de grandes quantités de données issues de plusieurs sources. L'objectif de l'ETL est de rendre les données hétérogènes compatibles avec la base de données de référence et l'organisation pour

une meilleure analyse. Pour y parvenir, l'ETL structure, nettoie, compile et agrège les données extraites de différents composants du système d'entreprise.

D'une manière générale, les outils ETL fonctionnent en trois étapes, à savoir :

- La première étape de l'ETL est l'extraction (*Extract*) des données à partir de sources de données diverses, telles que des bases de données opérationnelles, des fichiers plats, des applications tierces, etc. Ces données sont extraites dans leur format d'origine à travers des connecteurs adaptés. De plus, pour garantir l'actualisation des données, le processus d'extraction doit également inclure un mécanisme de synchronisation pour récupérer les données mises à jour depuis la dernière extraction.
- En effet, la phase de transformation des données est essentielle pour garantir la qualité et l'intégrité des données extraites. Cette étape implique des traitements tels que la vérification des données, la conversion de leur format, le nettoyage et la suppression des doublons. L'objectif principal est de s'assurer que les données soient cohérentes, fiables et exploitables pour la prochaine phase, à savoir le chargement dans l'entrepôt de données.
- Finalement, l'étape de chargement de l'ETL consiste à prendre les données transformées et nettoyées et à les charger dans l'emplacement cible qui est généralement un entrepôt de données ou un *datalake*. Cette étape peut également inclure la validation des données pour s'assurer qu'elles sont correctement chargées et qu'elles correspondent aux spécifications. Une fois les données chargées, elles sont disponibles pour être exploitées par les outils de BI et d'analyse de données pour aider à la prise de décision en entreprise.

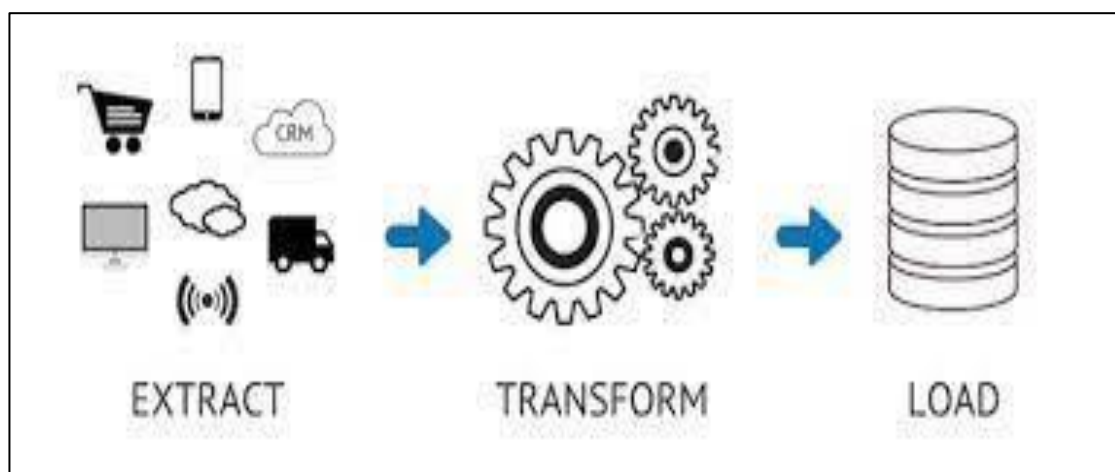


Figure 9: Processus ETL (Extract Transform Load)

2.4 Etude comparative des outils ETL

Effectivement, il existe de nombreuses solutions ETL sur le marché, chacune ayant ses propres fonctionnalités et capacités. Il est donc important de choisir un outil qui répond aux besoins et objectifs spécifiques de l'entreprise, en fonction de facteurs tels que le volume de données, les sources de données, les formats de données, les besoins en termes de temps de traitement et d'autres exigences spécifiques. Une évaluation complète des besoins de l'entreprise, des options disponibles sur le marché, ainsi que des coûts et des avantages associés à chaque solution doit être effectuée pour prendre la meilleure décision.

Le choix d'un ETL doit se faire en fonction de plusieurs critères à évaluer, à savoir :

- La portée du projet ETL ;
- La volumétrie des données.

Compte tenu de tous ces critères, nous avons choisi de faire l'étude comparative sur deux outils ETL, à savoir :

Talend : une entreprise spécialisée dans la gestion des données d'entreprise, fondée en 2005 à Suresnes. Elle propose une solution ETL appelée "*Talend Open Studio for Data Integration*", qui est une solution open source. Cette solution est considérée comme complète et permet de réaliser des projets ETL en un temps record.

Informatica est une entreprise éditrice de solutions d'intégration indépendante, considérée comme le leader mondial de la gestion des données dans le Cloud. Ses produits sont commercialisés sous licence propriétaire, ce qui implique que les utilisateurs doivent les acheter.

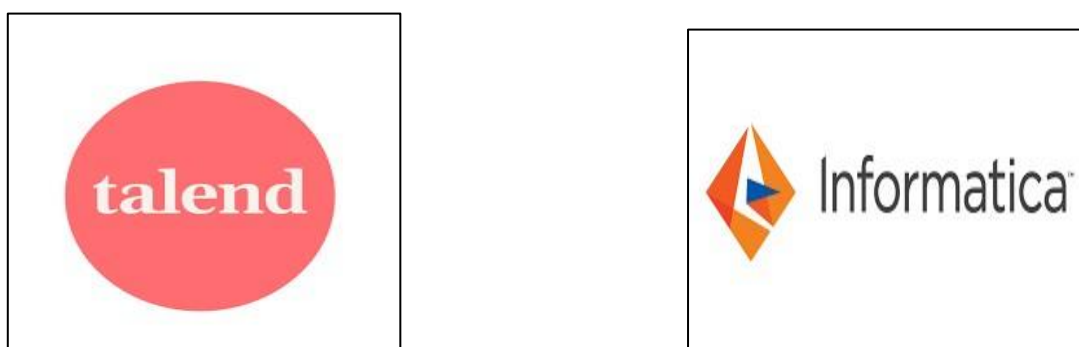


Figure 10: Logo Talend et Informatica

Caractéristiques	Talend	Informatica
Commercial ou open source	Implémente de nombreuses solutions d'intégration de données à la fois open source et éditions commerciales	Fournit uniquement l'intégration de données commerciales
Historique	Lancé octobre 2006	Crée en 1993
Tarification	Bénéficier gratuitement de l'édition open source	Licence payante
Popularité	Outil ETL open source le plus connu	Plateforme d'intégration de donnée d'entreprise mature et intégrée pour charges de travail ETL
Plateforme	Développe du code Java qui permet de fonctionner sur n'importe quelle plate-forme prenant en charge Java	Génère les métadonnées qui sont stockées dans le référentiel RDBMS, il ne développe aucun code
Codes personnalisés	Le code personnalisé peut être écrit précisément dans Talend	Intégration du code personnalisé par Java, conversion si efficace.
Déploiement	Facile à déployer	Besoin de faire progresser l'automatisation en matière de déploiement
Réutilisabilité	Les transformations peuvent être réutilisées dans Talend	Aide à la génération de composants réutilisables
Planification	Open source de Talend n'a pas de fonctionnalité planification mais le commercial oui.	Avec un gestionnaire de service dans Informatica, la planification des tâches se fait facilement
Parallélisme	Supporte le parallélisme avec éditions commerciales et non avec open source	Prend en charge le parallélisme, plusieurs sessions de mappage peuvent être réalisés sur le même serveur
Sauvegarde et restauration	Talend ne dispose d'aucune fonctionnalité de sauvegarde.	Le gestionnaire de référentiel aide aux plans de récupération et de sauvegarde.

Tableau 1: Etude Comparative de Talend vs Informatica

Après ces descriptions et les différentes caractéristiques de chaque outil, notre choix se porte sur

Talend Open Studio (TOS), puisqu'il est open source, et pour éditer des code Java et des codes personnalisés.

2.5 Modélisation multidimensionnelle

La modélisation dimensionnelle est une méthode de structuration des données qui diffère de la structure en troisième forme normale (3FN) souvent utilisée par les modélisateurs de systèmes de traitement transactionnel en ligne (en anglais On-Line Transactional Processing - OLTP). Cette méthode de modélisation aboutit à la création d'un modèle dimensionnel, qui constitue la base de la construction des cubes OLAP.

Nous distinguons principalement trois types de base de modèle dimensionnel :

- Le schéma en étoile
- Le schéma en flocons de neige
- Le schéma en Constellation.

2.5.1 Le schéma en étoile

Le modèle dimensionnel est une structure de données qui diffère de la structure en troisième forme normale (3FN) fréquemment utilisée pour les systèmes de traitement transactionnel en ligne (OLTP). Il se compose d'une table de faits centrale entourée de tables dimensionnelles qui contiennent des informations descriptives sur les faits. Lorsque représenté graphiquement, ce modèle ressemble à une étoile, d'où son nom de "modèle en étoile".

Voici la figure 11 illustrative du schéma en étoile :

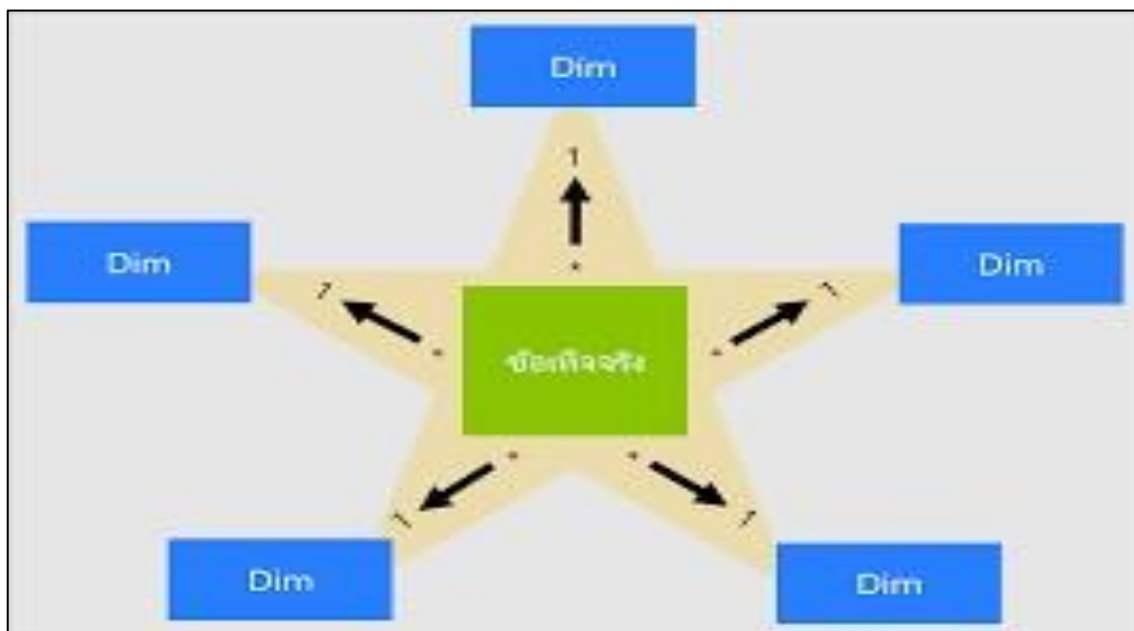


Figure 11: Schéma en étoile

2.5.2 Le schéma en flocons de neige

Le modèle en flocons de neige représente une variante du modèle en étoile qui se distingue par la normalisation des tables de dimension. Contrairement au modèle en étoile, les attributs de chaque niveau hiérarchique sont stockés dans des tables de dimensions distinctes. Bien que cette approche complexifie la structure des tables, elle facilite leur maintenance.

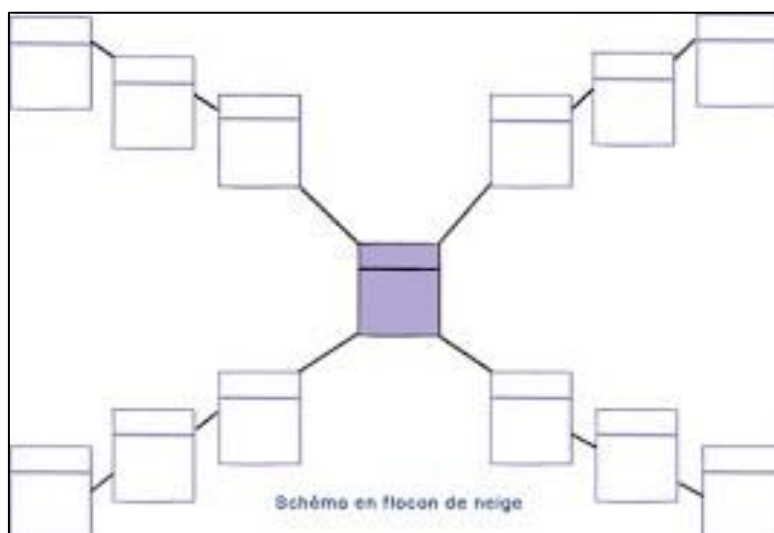


Figure 12: Schéma en flocon de neige

2.5.3 Le schéma en constellation

Le modèle en constellation est un des modèles de données multidimensionnelles utilisés pour concevoir des entrepôts de données. Il est composé de plusieurs tables de faits qui partagent des tables de dimensions communes. Contrairement au modèle en étoile, il peut être vu comme une collection de plusieurs schémas d'étoiles interconnectés. En raison de sa complexité, la modélisation des relations entre les tables doit être soigneusement pensée pour assurer des requêtes efficaces et des performances optimales.

Pour les systèmes complexes, il est recommandé d'utiliser le schéma en constellation.

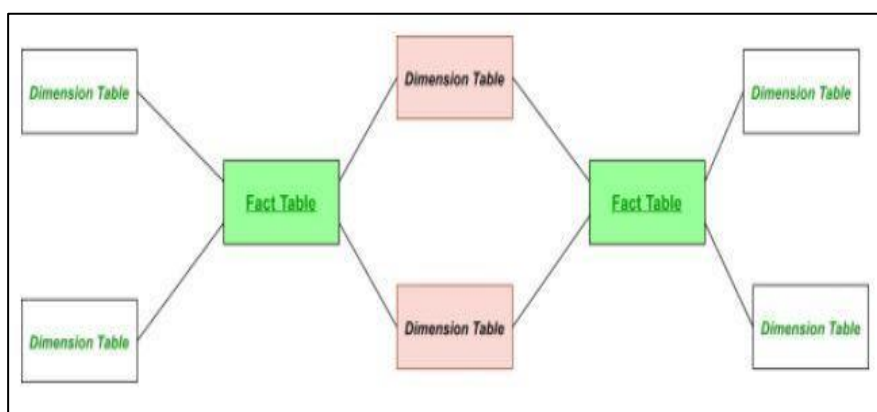


Figure 13: Schéma en constellation

2.5.4 La table de « Faits »

Dans un entrepôt de données, une table de faits est une table qui contient des données quantitatives sur un sujet à étudier en fonction de divers axes d'analyse appelés dimensions. Ces tables représentent des associations qui dépendent de l'existence des occurrences correspondantes dans les tables dimensionnelles. Les mesures de la table de faits reflètent les informations de l'activité à l'analyse. Cependant, il est important de souligner que certaines tables de faits peuvent être dépourvues d'attributs et ne servir qu'à relier les tables dimensionnelles entre elles.

La table de faits a les caractéristiques suivantes :

- Une table de fait contient les valeurs numériques de ce qu'on désire mesurer
- Une table de fait contient les clés associées aux associés aux dimensions. Il s'agit clés étrangères dans la table de faits
- En générale une table de faits, un petit nombre de colonnes
- Une table de faits contient plus d'enregistrement qu'une table de dimension
- Toutes les colonnes représentant les faits (mesure1, mesure2, etc.) dans la table de fait doivent référer et avoir un lien direct aux clés de dimensions

Une table de fait se présente de la manière suivante :

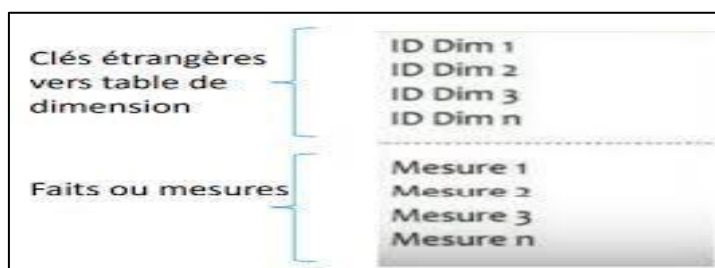


Figure 14: Illustration d'une table de Faits

2.5.5 La table de dimension

Une table de dimension est une table de base de données qui contient des attributs permettant de classer ou décrire les enregistrements dans une table de faits. Chaque table de dimension est liée aux enregistrements correspondants dans la table de faits via une clé primaire. Cette clé primaire est souvent utilisée comme clé étrangère dans la table de faits pour associer les informations provenant des différentes tables de dimension aux données numériques de la table de faits.

2.6 Analyse

2.6.1 On-Line Analytical Processing (OLAP)

L'OLAP (Online Analytical Processing) est un processus permettant d'analyser des données à différents niveaux de détail (par exemple année, trimestre, mois, pays, département) et selon différentes dimensions (telles que le temps, le client, le produit, la localisation géographique). Selon Ralph Kimball, expert en informatique, l'OLAP consiste à interroger et à présenter de manière globale des données numériques et textuelles stockées dans un entrepôt de données, en utilisant une méthode d'interrogation spécifiquement dimensionnelle.

2.6.1.1 Fonctionnement de OLAP

L'Online Analytical Processing (OLAP) permet d'analyser les données à différents niveaux de détails et dimensions, tels que le temps, le client, le produit, etc. Pour cela, les données sont collectées à partir de diverses sources et stockées dans un entrepôt de données. Ensuite, elles sont nettoyées, organisées en cubes de données et classées par dimensions, qui sont constituées de membres organisés de manière hiérarchique. Les cubes OLAP sont souvent pré-résumés dans toutes les dimensions pour améliorer la vitesse de requête par rapport aux bases de données relationnelles.

2.6.1.2 Les opérations d'analyse OLAP

Les systèmes OLAP ont pour objectif de permettre l'analyse de données à travers plusieurs dimensions. Les analystes peuvent ainsi effectuer cinq types d'opérations d'analyse OLAP sur ces bases de données multidimensionnelles.

Le premier type d'opération OLAP est appelé "Slice". Il permet de sélectionner une seule valeur pour une dimension spécifique afin de visualiser les données correspondantes.

Le deuxième type d'opération OLAP est appelé "Dice". Il permet de sélectionner des valeurs pour plusieurs dimensions simultanément.

Le troisième type d'opération OLAP est appelé "Roll-up" ou "Drill-up". Il permet de résumer les données en remontant d'un niveau hiérarchique à un autre niveau supérieur dans une dimension.

Le quatrième type d'opération OLAP est appelé "Drill-down" ou "Drill-in". Il permet de détailler les données en descendant à un niveau hiérarchique inférieur dans une dimension.

Enfin, le cinquième type d'opération OLAP est appelé "Pivot". Il permet de basculer les axes de deux dimensions différentes pour visualiser les données sous un autre angle. Grâce à ces opérations, les analystes peuvent explorer et analyser les données sous différents angles et niveaux de détails, ce qui leur permet de prendre des décisions éclairées.

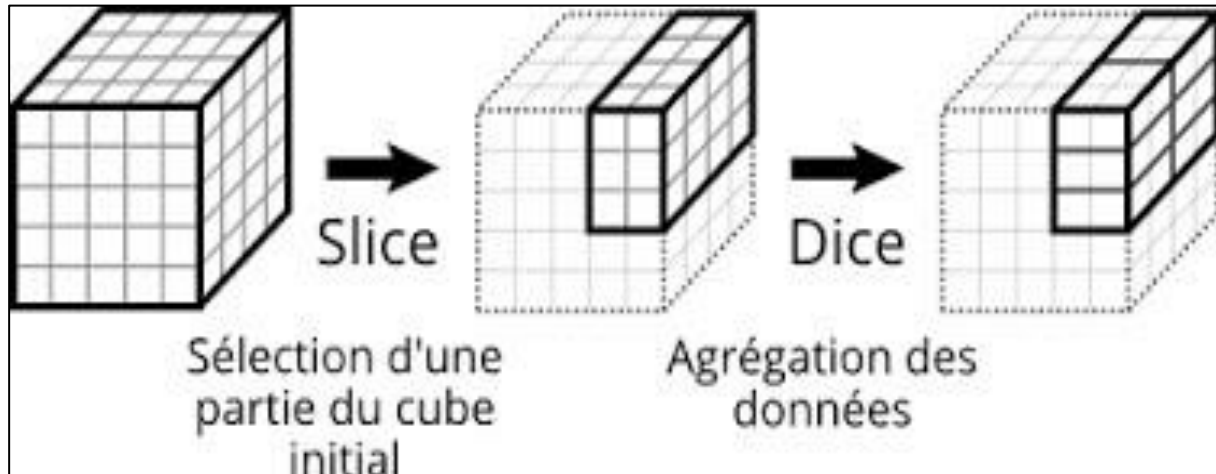


Figure 15: Différence entre Slice et Dice

2.7 La visualisation

Effectivement, la visualisation de données est un processus clé pour comprendre et analyser les données. Elle permet de présenter les données sous forme graphique ou visuelle, facilitant ainsi la compréhension et l'interprétation des informations. La visualisation de données permet de regrouper les données en tableaux, graphiques, cartes ou infographies, qui sont ensuite intégrés dans un tableau de bord. Cela permet aux entreprises de mieux comprendre les données qu'elles ont collectées, et de prendre des décisions plus éclairées en temps réel. En résumé, la visualisation de données est un outil important pour les entreprises qui cherchent à exploiter leurs données de manière efficace, pour améliorer leur agilité opérationnelle et leur compétitivité.

2.7.1 Le reporting

Le reporting est une méthode de la Business Intelligence qui implique la création et la distribution de rapports d'activité préétablis à partir de données collectées dans des sources de données. Cette méthode permet de synthétiser les données de manière claire et concise pour faciliter la prise de décision et le suivi des indicateurs clés de performance (KPI). Les rapports sont généralement créés à partir de requêtes SQL préétablies et sont conçus pour répondre à des questions spécifiques de l'entreprise. Les outils de reporting offrent également des fonctionnalités pour la création de modèles de rapports, l'analyse de données, la conception de graphiques et la mise en page pour rendre les rapports plus attrayants et utiles.

2.7.2 Tableau de bord (ou Dashboard)

En effet, le tableau de bord est un outil de pilotage stratégique qui permet de suivre en temps réel les indicateurs clés de performance (KPI) de l'entreprise et de mesurer l'atteinte des objectifs fixés. Il peut être utilisé à différents niveaux de l'organisation, du niveau opérationnel au niveau stratégique. Le tableau de bord peut être constitué de différents types de graphiques, tels que des jauges, des camemberts, des histogrammes ou des diagrammes en barres, pour faciliter la lecture des indicateurs. Il est également souvent accompagné de commentaires ou de notes pour expliquer les tendances et les écarts constatés. Le tableau de bord est donc un outil essentiel pour la prise de décision, la gestion de projet et la communication interne de l'entreprise.

Un tableau bord doit :

- Permettre aux décideurs d'identifier les écarts le plus rapidement possible et d'effectuer des actions correctives
- Être un outil de communication en interne
- Être également un outil de motivation au sein de l'entreprise, en mettant en lumière des objectifs de l'entreprise et sa stratégie
- Favoriser l'apprentissage continu en recherchant constamment à améliorer la performance de l'entreprise.

Exactement, le tableau de bord est plus orienté vers une vision synthétique et globale de la performance de l'entreprise, tandis que le reporting permet d'obtenir des informations plus détaillées et spécifiques sur une activité ou un processus en particulier. Le tableau de bord est donc un outil de pilotage stratégique, alors que le reporting est plutôt un outil de suivi opérationnel. En combinant ces deux outils, les entreprises peuvent avoir une vue d'ensemble complète de leur activité et des informations pertinentes pour prendre des décisions éclairées et améliorer leur performance globale.

2.8 Déploiement

Les pipelines de déploiement offrent une manière cohérente et contrôlée de gérer le cycle de vie de ces différents types de contenu dans l'environnement Power BI. Ils permettent notamment de faciliter le déploiement et la mise à jour de ces éléments, de réduire les risques d'erreur humaine ou de conflits entre différentes versions, et de garantir la cohérence et la qualité du contenu publié pour les consommateurs finaux. Les pipelines de déploiement peuvent également être configurés pour automatiser les tâches répétitives et réduire les délais de mise sur le marché.

Conclusion

Ce chapitre a abordé les notions fondamentales de l'informatique décisionnelle ainsi que les outils nécessaires pour mettre en place une solution BI. Dans le prochain chapitre, nous allons nous concentrer sur l'identification des besoins fonctionnels et non fonctionnels pour élaborer une solution BI adaptée.

Chapitre 3 : Sprint 0 - Analyse et spécification des besoins

Introduction

Dans ce chapitre, nous abordons l'étape cruciale de l'analyse des besoins, qui implique l'identification des différents acteurs impliqués dans l'application, ainsi que la collecte et l'analyse des besoins fonctionnels et non fonctionnels. Nous présentons également le Backlog de produit, qui est une liste exhaustive des fonctionnalités à développer pour répondre aux besoins des utilisateurs, et la planification des sprints, qui permet de déterminer les tâches à réaliser pour chaque itération du projet.

1. Spécification des besoins

Dans cette section, nous allons procéder à l'identification des acteurs impliqués dans notre système et présenter les besoins du projet.

1.1 Identification d'acteurs

Notre système retient un seul acteur à savoir :

- Utilisateur (décideur) : personne physique ou morale qui utilise le système pour prendre des décisions relatives au financement ou à la réalisation d'un projet. Dans le contexte de la BI, le décideur est responsable de la lecture, de l'évaluation et de l'interprétation du tableau de bord pour prendre des décisions éclairées.

1.2 Analyse des besoins fonctionnels

1.2.1 Besoins fonctionnels

Les besoins fonctionnels de notre système sont les suivants :

- Collecte et stockage des données : notre système doit être capable de collecter et stocker les données relatives aux activités commerciales de la société WorldSoft pendant une période donnée.
- Transformation et chargement des données : les données collectées doivent être transformées et chargées dans l'entrepôt de données de manière à être prêtes pour l'analyse.
- Analyse des données : notre système doit être capable de réaliser des analyses des données stockées dans l'entrepôt de données pour fournir des informations utiles et pertinentes aux décideurs.

- Visualisation des données : les résultats de l'analyse doivent être présentés sous forme de rapports et de tableaux de bord pour faciliter la prise de décision.

Les besoins de l'entrepôt de données

La conception de l'entrepôt de données (Data Warehouse) vise plusieurs objectifs, notamment :

- Créer une source unique et stable de données, qui ne sera pas modifiée lors des opérations d'analyse ou de requête
- Assurer un meilleur contrôle des données en les organisant selon les besoins de l'entreprise
- Effectuer une modélisation multidimensionnelle des données, qui permet de visualiser les données sous forme de cubes, de tableaux de bord et de graphiques
- Assurer la flexibilité et la scalabilité de la base de données, de sorte qu'elle puisse être adaptée aux besoins futurs de l'entreprise
- Assurer la traçabilité des données, c'est-à-dire conserver l'historique et la source des données pour garantir leur fiabilité et leur qualité.

Les besoins de l'application

- Disponibilité des données à n'importe quel moment
- Interroger les données de l'entrepôt de données (Data Warehouse)
- Visualiser la statistique :
 - Vente de produit (Ticket) par catégorie du client
 - Vente de produit par produit
 - Vente de produit par client
 - Vente de produit par catégorie client
 - Chiffre d'affaires par clients , pays et entité
 - Profit par clients ,pays et entité

1.2.2 Besoins non fonctionnels

Les besoins non fonctionnels sont aussi importants dans la conception de notre système. Voici quelques exemples de besoins non fonctionnels pour notre projet :

- Sécurité : notre système doit garantir la sécurité et la confidentialité des données stockées et analysées.
- Performance : notre système doit être capable de traiter de grandes quantités de données de manière efficace et efficiente.

- Disponibilité : notre système doit être disponible en tout temps pour permettre aux décideurs d'accéder aux informations nécessaires pour prendre des décisions.
- Évolutivité : notre système doit être facilement évolutif pour répondre aux besoins futurs de la société WorldSoft.

Après cette identification des besoins, en passant par l'identification des acteurs, nous allons passer par la planification du projet.

2. Pilotage du projet par Scrum

2.1 Identification de l'équipe Scrum



Figure 16: Présentation de l'équipe Scrum

Notre équipe Scrum est composée de trois personnes, à savoir :

- Product Owner qui est représenté par Monsieur Wassim
- Scrum Master sera Madame Fatma Kaabi
- Et enfin l'équipe de développement est constituée par Zaki Bouzidi

2.2 Elaboration de Product Backlog

Le *Product backlog* est un élément clé de la méthodologie Agile et permet de définir les fonctionnalités à implémenter dans l'ordre de priorité pour répondre aux besoins des utilisateurs. Il contient une liste de "*User Stories*" ou "Récits Utilisateurs" qui sont classées par ordre de priorité. Chaque élément de la liste est associé à un identifiant unique (ID), une fonctionnalité, une description et une priorité.

L'identifiant permet de différencier chaque élément de la liste et de le suivre tout au long du processus de développement. La fonctionnalité décrit l'objectif de l'élément et ce qu'il doit être capable de faire. La description fournit plus de détails sur les exigences et les spécifications du produit. La priorité indique l'importance de l'élément dans la liste et permet de déterminer l'ordre dans lequel les fonctionnalités seront développées.

Le *Product backlog* est constamment mis à jour au fil du temps en fonction des besoins et des changements apportés au projet. Il est utilisé pour guider les itérations de développement et pour s'assurer que les fonctionnalités les plus importantes sont développées en premier.

Le tableau ci-dessous résume le *backlog* produit de notre système.

ID	Fonctionnalité	Description	Priorité
1	Consulter le tableau de bords	L'utilisateur ne peut que consulter le tableau de bord	Élevé
2	Consulter les rapports	Etant utilisateur je peux seulement les rapports	Elevé

Tableau 2: Tableau du Product backlog

2.2.1 Diagramme de cas d'utilisation

Par définition, un diagramme de cas d'utilisation est un diagramme UML utilisé pour donner une vision globale du comportement fonctionnel d'un système logiciel. Un cas d'utilisation représente une unicité discrète d'interaction entre un utilisateur et un système. Alors le diagramme de cas d'utilisation de notre système se présente ainsi :

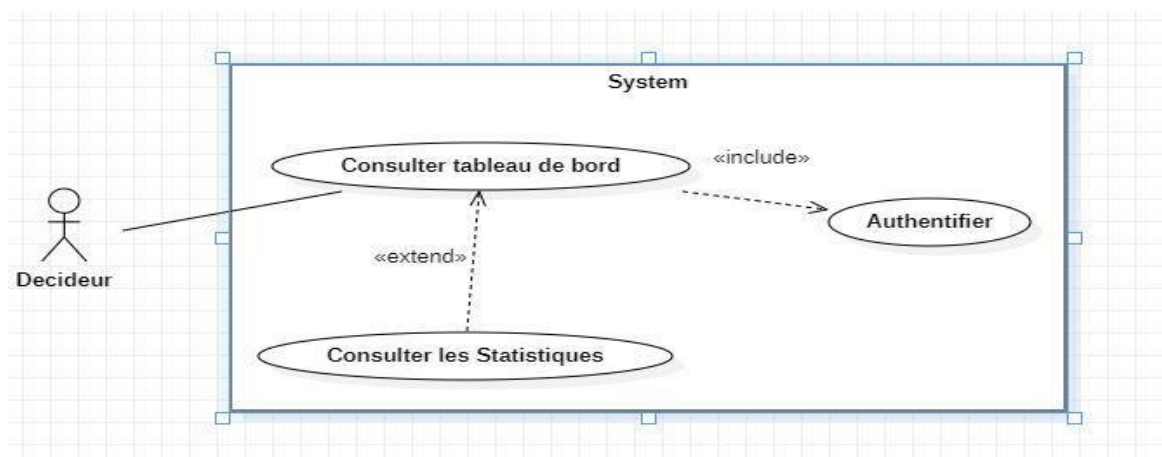


Figure 17: Diagramme de cas d'utilisation

2.3 Découpage du projet

La réunion de planification de release est considérée comme l'un des événements clés dans la méthode Scrum. Son objectif principal est de préparer un plan de travail pour le projet en identifiant les tâches prioritaires à effectuer et en déterminant le backlog des sprints.

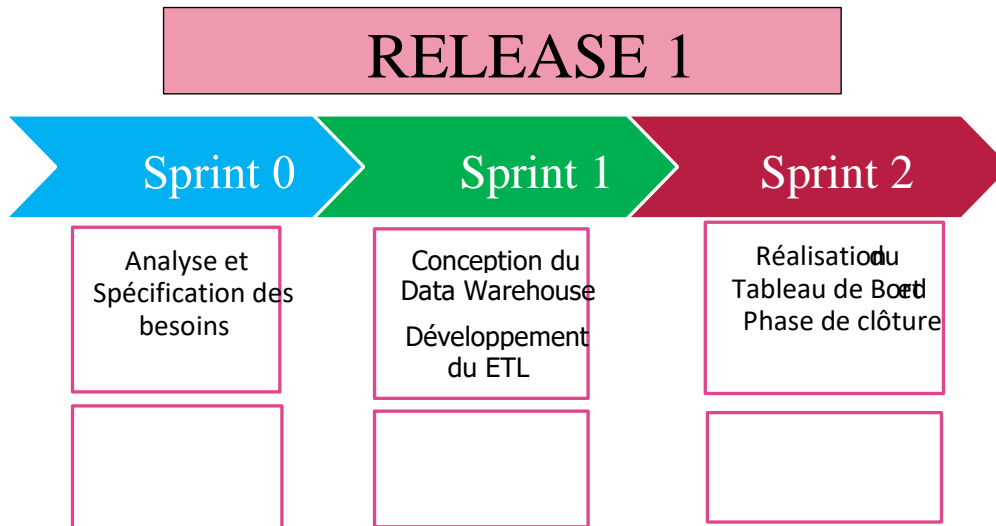


Figure 18: Découpage du projet

Conclusion

Dans ce chapitre, nous avons clairement défini les besoins du projet en identifiant les acteurs, les besoins fonctionnels et non fonctionnels. Nous avons également présenté les membres de l'équipe qui ont contribué à l'élaboration du projet et avons divisé le projet en différentes parties. Dans le prochain chapitre, nous allons entrer dans les détails des différents sprints.

Chapitre 4: Sprint 1 – Conception du Data Warehouse et Développement du ETL

Introduction

Le Sprint est une période d'itérations successives qui se terminent lorsque les incréments atteignent un produit fonctionnel. Pour cette phase, notre attention se porte sur la solution Business Intelligence (BI), la conception et la création du Data Warehouse (Entrepôt de données) et la mise en place du processus Extract Transform Load (ETL) en utilisant l'outil Talend Open Studio for Data Integration (TOS).

1. Conception du Data Warehouse

Pour élaborer un système de Business Intelligence (BI) efficace, la modélisation est une étape cruciale. Cette phase permet de concevoir un modèle dimensionnel qui prend en compte les besoins des utilisateurs et les objectifs du projet. Pour cela, il est nécessaire de suivre un processus en trois étapes qui consistent à sélectionner les dimensions, les mesures et le modèle dimensionnel approprié. Ce choix doit se baser sur une compréhension approfondie des données à traiter et des besoins des utilisateurs pour garantir une conception optimale du système BI.

1.1 Choix des dimensions

Une table de dimension est une table dans une base de données qui fait référence à des informations ou attributs définissant des enregistrements spécifiques dans la table de base de données principale. Les tables de dimension contiennent des descriptions textuelles des activités et sont destinées à être utilisées en conjonction avec les tables de faits. Elles facilitent les analyses en tranches et en dés.

1.1.1 Dim_Airline

Cette dimension étudie la description des compagnies et leurs destinations.

Désignation	Détails	Type de données
Id_Comp	La clé primaire de la compagnie	Number
Code_IATA	Code Compagnie	Chaine de caractère
Nom_Comp	Nom de la compagnie	Chaine de caractère
Domestic	Vol intérieur du pays ou pas	Caractère

Tableau 3: Table de dimension Airlines

1.1.2 Dim_Customer (Dim_Client)

Dans une entreprise, la satisfaction du client est un enjeu majeur pour réaliser des bénéfices. Il est donc important de se différencier de la concurrence en répondant au mieux aux attentes du marché et de ses clients.

La dimension client représente les caractéristiques des clients, présentées dans le tableau ci-dessous:

Désignation	Détails	Type de données
Id_Customer	Clé Primaire du client	Number
Id_Entité	Clé primaire de l'entité devenu clé étrangère du client	Number
Id_Pays	Clé primaire du pays devenu clé étrangère du client	Number
Id_Cat_Cust	Clé primaire de la catégorie client devenu clé étrangère du client	Number
Cust_Name	Nom du client	Chaine de caractère
Nom_pays	Nom du pays	Chaine de caractère
Nom_cat_cust	Nom catégorie du client	Chaine de caractère

Tableau 4: Table de dimension Client

1.1.3 Dim_Country (Dim_Pays)

Puisque la société "WorldSoft" travaille dans le domaine de la réservation de billets en ligne, il est important de disposer d'une table de dimension pays pour connaître l'origine de chaque réservation ainsi que la destination.

Cette dimension est décrite dans le tableau ci-dessous :

Désignation	Détails	Type de données
Id_Pays	Clé primaire de la table pays	Number
Nom_Pays	Nom du pays	Chaine de caractère

Tableau 5: Table de dimension Pays (Country)

1.1.4 Dim_Entité

La table de dimension "Dim_Entité" représente les agences de voyages partenaires de la société "WorldSoft" et dont les transactions sont enregistrées dans sa base de données.

Les informations de cette dimension sont décrites dans le tableau suivant :

Désignation	Détails	Type de données
Id_Entité	Clé primaire de la table Entité	Number
Id_Pays	Clé étranger	Number
Ent_Name	Nom de l'entité ou agence	Chaine de Caractère
Ent_Ville	Ville de se situe l'entité	Chaine de Caractère

Tableau 6: Table de dimension entité

1.1.5 Dim_Cust_Cat (Catégorie Client)

La table de dimension de catégorie de client permet de classer les clients en deux catégories distinctes, à savoir les clients publics et les clients privés. Cette classification permet une meilleure organisation et structuration de l'ensemble de la clientèle de la société.

Les détails de cette dimension sont présentés dans le tableau suivant :

Désignation	Détails	Type de données
Id_Cust_Cat	Clé primaire de la table	Number
Name_Cust_Cat	Nom de catégorie du client	Chaine de Caractère

Tableau 7: Table de dimension Catégorie client

1.1.6 Dim_GDS (Global Distribution System)

Les GDS (Global Distribution System) sont des plateformes de distribution de services touristiques tels que des offres de vols, d'hôtels, de locations de voitures, de trajets en train, etc. pour les intermédiaires tels que les agences de voyages, les tour-opérateurs et les pure-players. La table de dimension "Dim_GDS" a pour but de décrire les informations liées aux différents GDS utilisés par la société "WorldSoft".

Le tableau ci-dessous illustre les attributs de cette table :

Désignation	Détails	Type de données
Id_GDS	Clé primaire de la table GDS	Number
Id_Entité	Clé étrangère de la table	Number
Name_GDS	Nom du Système Globale de Distribution (GDS)	Chaine de caractère

Tableau 8: Table de dimension Global Distribution System (GDS)

1.2 Le choix des mesures

Le « Key Performance Indicator » (K.P.I) ou Clé d'indicateur de performance est une information ou un ensemble d'informations permettant et facilitant l'appréciation, par un décideur, d'une situation donnée.

Une mesure appartient à un fait, et ce fait possède certaine granularité. Les mesures ont deux propriétés suivantes :

- Valeur numérique
- Les fonctions pouvant dériver la mesure à partir d'autre mesure.

Les mesures de notre table de fait se présentent dans le tableau suivant :

Indicateurs	Details	Formules
SUM_QTY_TKTT	Quantité de ticket vendu	COUNT(ETAT_TICKET)
GROSS_SALES	Vente brute	SUM(MONTANT_A_FACTURER)
GROSS_COST	Coût brut	SUM(MONTANT_DEV_SOURCE)
GROSS_PROFIT	Profit brut	SUM (GROSS_SALES – GROSS_COST)

Tableau 9: Table des Mesures

1.3 Table de fait « Fact_Sales »

Une table est une table qui contient les données observables (les faits) que l'on possède sur un sujet et que l'on veut étudier, selon divers axes d'analyses (dimensions). Les faits ou mesures dans un entrepôt de données sont généralement numériques, puisque d'ordre quantitatif.

La table de fait contient les faits où les mesures de performances sont stockées, les clés étrangères qui sont les clés primaires des tables de dimensions.

La structure de notre table de faits se présente comme suit :

Désignation	Détails	Type de données
Id_Entité	Clé étrangère	Number
Id_Cust	Clé étrangère	Number
Id_GDS	Clé étrangère	Number
Id_Comp	Clé étrangère	Number
Time_Code	Clé étrangère	Number
Code_IATA	Code avion attribué par IATA	Chaine de caractère
Name_GDS	Nom du GDS Utilisé	Chaine de caractère
Domestic	Vol intérieur du pays ou pas	Caractère
Cust_Name	Nom du client	Chaine de caractère
Ent_Name	Nom de l'entité	Chaine de caractère
Ent_ville	Ville de client	Chaine de caractère
SALES	Vente brute	Number
PROFIT	Profit brut	Number
TYPE_DEV	Devise utilisée de vente	Chaine de caractère

Tableau 10: Description de la table de Fait

1.4 Structuration du Data Warehouse

Nous avons opté pour le modèle en étoile pour la conception de notre data warehouse, car il convient parfaitement à la modélisation de notre entrepôt de données. L'entrepôt de données est constitué d'une table de faits et de six tables de dimensions. La table de faits comporte plusieurs clés primaires et étrangères qui correspondent aux clés des tables de dimensions, ainsi que des indicateurs qui serviront à la création du tableau de bord.

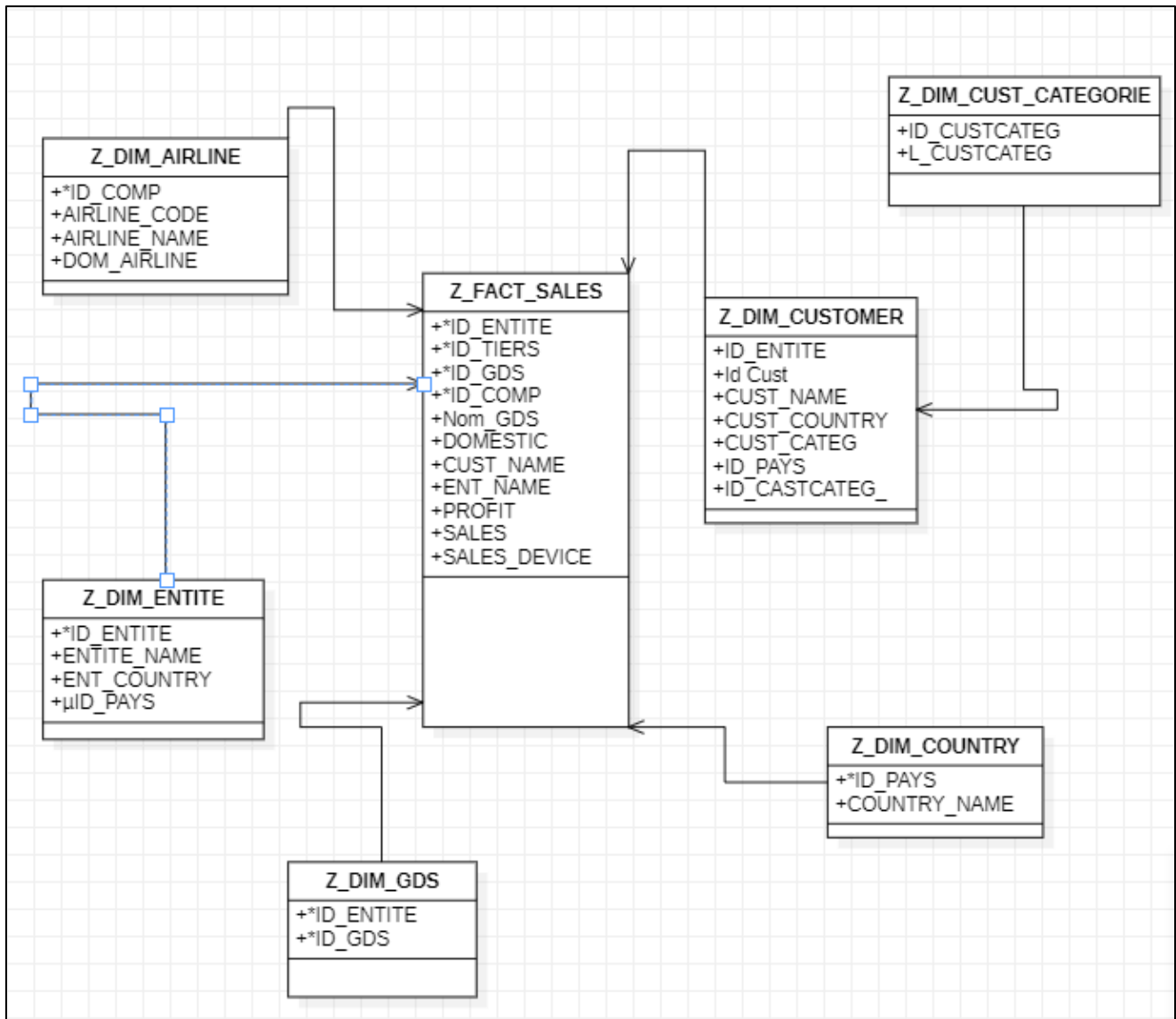


Figure 19: Modélisation de l'entrepôt de données

2. Conception de l'ETL (Extract – Transform – Load)

Une fois que le Data Warehouse ou entrepôt de données est conçu, il est temps de le remplir de données en utilisant le processus ETL.

Dans le cas de la société "WorldSoft", qui utilise un Système de Gestion de base de données (SGBD) Oracle, l'alimentation de l'entrepôt de données se déroulera en quatre phases.

- Tout d'abord, nous examinerons la base de données sources pour identifier les données pertinentes pour notre entrepôt.
- Ensuite, nous extrairons ces données.
- Dans la troisième phase, nous transformerons et traiterons les données extraites.
- Enfin, nous chargerons les données transformées dans le Data Warehouse.

En suivant ces quatre étapes clés, nous pouvons garantir une alimentation optimale de l'entrepôt de données.

2.1 Etude de données

Dans cette section, nous allons étudier les données de la source pour identifier les tables pertinentes à extraire. Nous allons ensuite alimenter les tables de dimension en utilisant ces données et concevoir le Data Warehouse. Pour commencer, nous devons nous connecter à la base de données locale.

2.1.1 Base de données

2.1.1.1 Connexion à la base de données

La base de données de la société est basée sur le SGBD Oracle (SQL). Pour nous connecter à cette base de données, nous utilisons le logiciel Oracle SQL Developer. La connexion à la base de données est illustrée dans la figure 20. Une fois que nous sommes connectés à la base de données de la société, nous créerons les tables de dimension et la table de faits en utilisant des requêtes SQL.

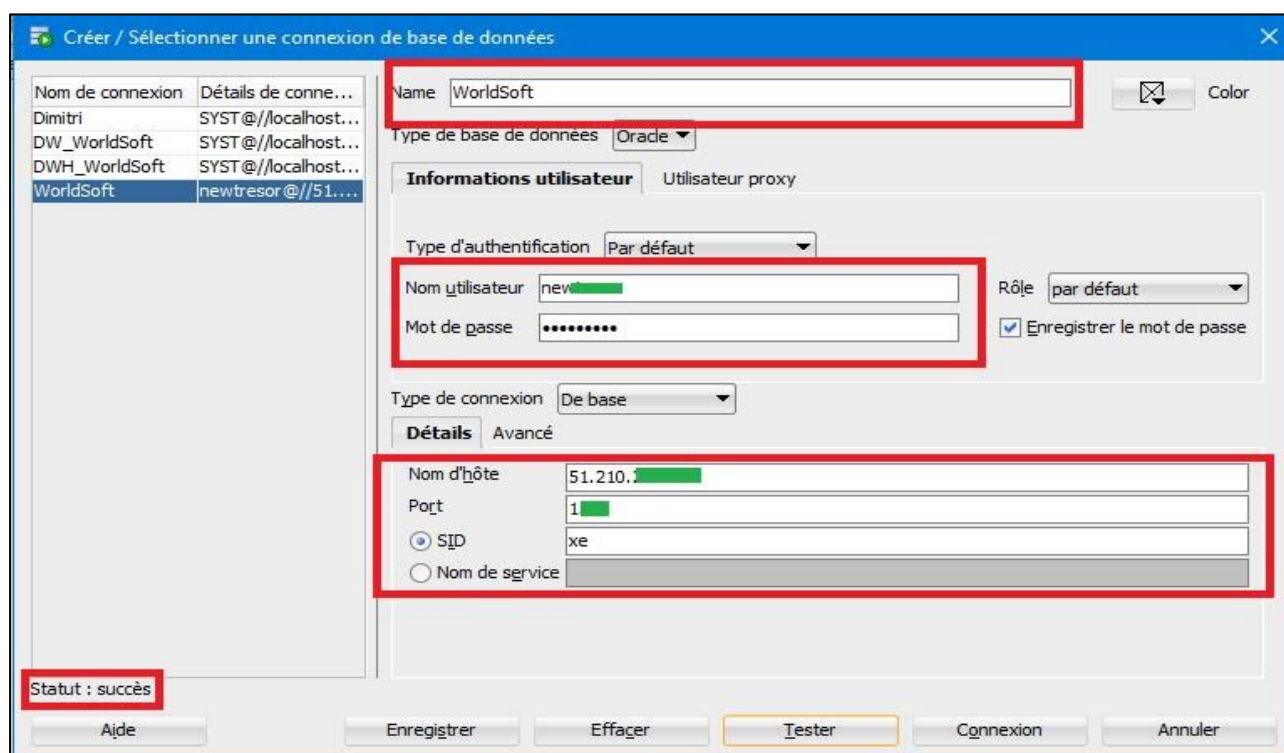


Figure 20: Interface de connexion à la base de données

2.2 Réalisation de l'ETL

Dans le deuxième chapitre, nous avons mentionné que pour réaliser le processus ETL, nous avons choisi d'utiliser l'outil Talend Open Studio. Il est important de rappeler que le processus ETL est un processus automatisé qui permet de prendre des données brutes, d'extraire les informations nécessaires à l'analyse, de les transformer en un format répondant aux besoins opérationnels, et de les charger dans un Data Warehouse.

2.2.1 Connexion de Talend à la SQL Developer

Afin de réaliser le processus ETL, notre première étape consiste à connecter l'outil ETL que nous avons choisi à la base de données pour récupérer les schémas ou les tables nécessaires pour remplir le Data Warehouse. L'interface de connexion à Talend est illustrée dans la figure 21.

The screenshot shows the 'Connexion base de données' dialog box in Talend. The title bar reads 'Connexion base de données'. The main title is 'Mise à jour de la connexion à la base de données - Étape 1/2'. Below the title, it says 'Mettre à jour les propriétés'. The dialog has several input fields: 'Nom' (containing 'WorldSoft_Groupe'), 'Objectif', 'Description', 'Créé par' (containing 'user@talend.com'), 'Verrouillé par' (containing 'user@talend.com'), 'Version' (containing '0.1'), 'Statut', and 'Chemin d'accès'. At the bottom, there are four buttons: '< Back', 'Next >', 'Finish', and 'Cancel'. The 'Next >' button is highlighted with a red box, and a red arrow points down to it.

Figure 21: Interface 1 de la connexion de Talend

Après cliquer sur « NEXT », Talend nous affiche l'interface ci-dessous (figure 22). Quand nous cliquons sur « Tester la connexion », il y'a une boîte de dialogue qui nous affiche le message.

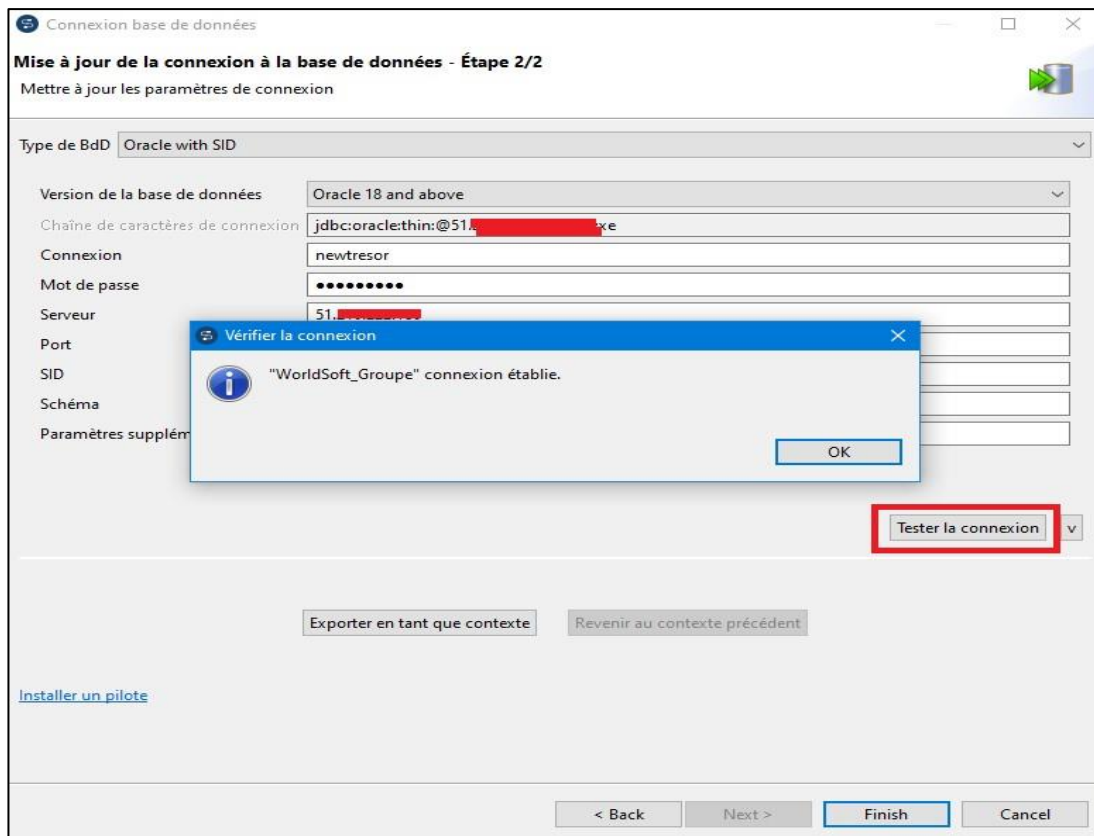


Figure 22: Connexion entre Talend et SQL Developer

2.2.2 Etape 1 : Récupération des Schémas

Après avoir établie la connexion entre Talend et la base données, nous allons tout d'abord récupérer les tables.

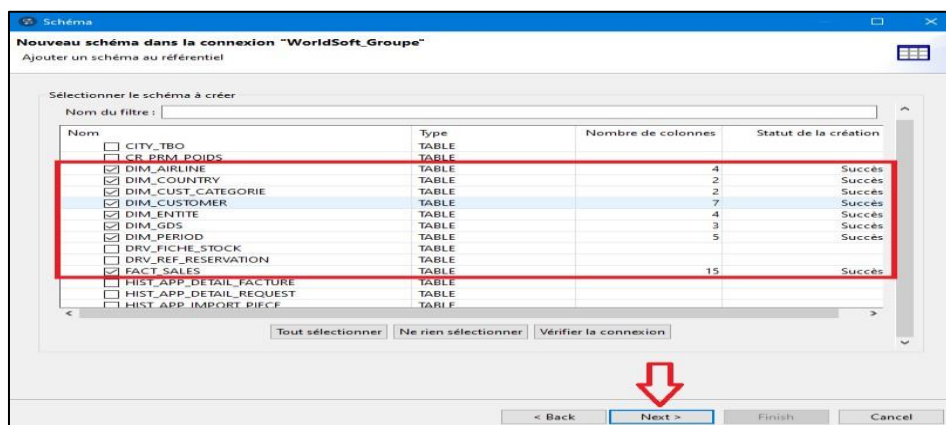


Figure 23: Récupération des tables

Avant de cliquer sur « Next », on doit dans un premier temps sélectionner les tables qui serviront de remplir le Data Warehouse.

2.2.3 Création des Jobs

Un Job est un ensemble de tâches regroupées en une unité fonctionnelle, qui consiste en plusieurs composants interdépendants. Il peut être exécuté en mode batch ou interactif, avec des entrées/sorties

multiples au sein d'un même job. Dans le cadre de notre projet, nous créerons des jobs qui seront exécutés manuellement en utilisant les composants correspondants.

La figure ci-dessous nous illustre le processus de création d'un job.

Figure 24: Création d'un Job

Après avoir terminé la création du job, nous pouvons voir notre job nouvellement créé dans la liste de Job ci-dessous.

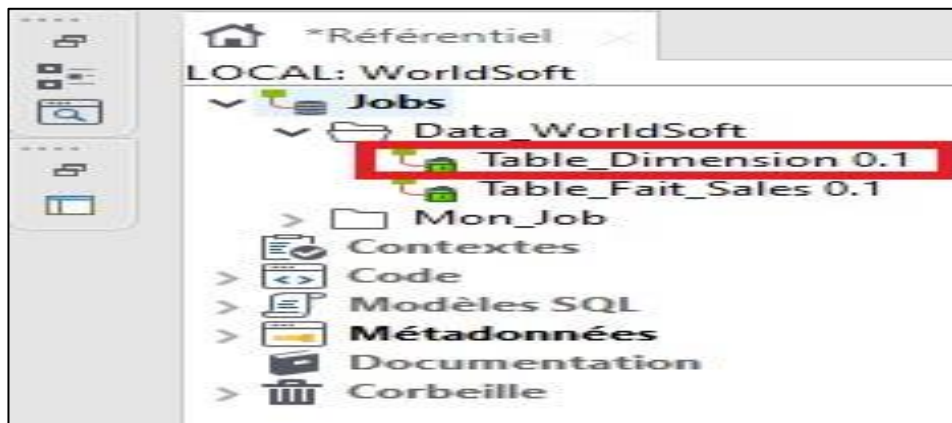


Figure 25: Job créée et opérationnel

Les jobs créés nous permettent de remplir les tables de dimensions et aussi la table de faits (Fact_Sales) et de les exécuter au fur et à mesure.

Pour cela, nous allons ajouter les composants oracle (inputs/output) et les composants Talend (filtre).

La figure 25 ci-dessous illustre notre job qui permet de remplir les tables de dimensions.

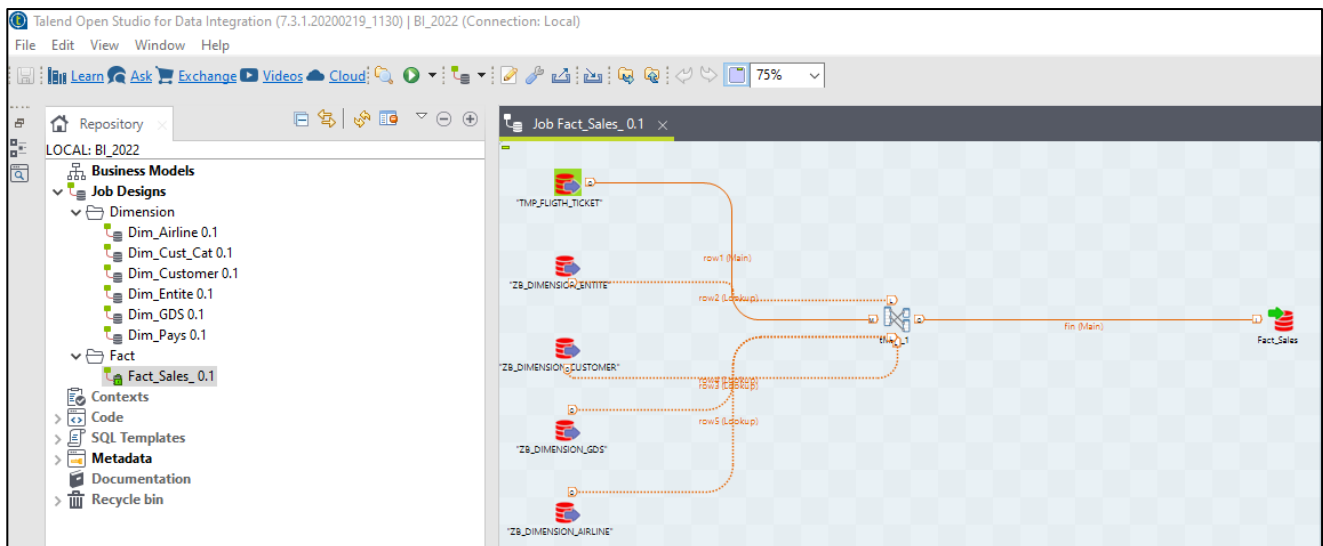


Figure 26: Remplissage des tables de dimension

Conclusion

Au cours de ce chapitre, nous avons mis en place le processus de conception et de remplissage de l'entrepôt de données en utilisant l'outil ETL. Cette étape est essentielle pour permettre l'exploitation des données et la réalisation de notre tableau de bord, qui constituera la deuxième partie de notre travail.

Chapitre 5: Sprint 2 – Réalisation du Tableau de bord et Phase de clôture

Introduction

Après avoir fini la partie conception de notre application, nous aborderons dans ce chapitre la partie réalisation de notre projet. La phase de réalisation est considérée comme étant la concrétisation finale de toute la méthode de conception. Alors pour cela, nous allons tout d'abord commencer par la présentation de notre environnement de travail à savoir l'environnement matériel et logiciels, ensuite nous présenterons l'outil utilisé pour la mise réalisation du tableau de bord et enfin nous allons conclure le chapitre par la présentation de quelques interfaces réalisées.

1. Environnement de travail

1.1 Environnement matériel

L'environnement matériel désigne la partie physique à l'aide de laquelle notre application est mise en place. Pour cela, nous utilisons un ordinateur portable dont les caractéristiques sont les suivantes :

- Ordinateurs Portables
 - Marque : HP EliteBook
 - Système d'application : Windows 10 Professionnel 64 bit
 - Processeur : Core i5 vPro
 - Mémoire RAM : 16Go
 - Disque Dur : 250 Go

2. Environnement logiciel et langage de développement

2.1 Environnement logiciel

L'environnement logiciel désigne l'ensemble des outils utilisés pour l'implémentation d'une solution informatique. Alors pour notre solution, nous avons décidé d'utiliser les outils suivants :

Nom	Description	Logo
SQL Developer	Un environnement de développement intégré gratuit destiné à faciliter la gestion et le développement des bases de données Oracle, que ce soit dans des déploiements traditionnels ou dans le cloud. Il est compatible avec les systèmes d'exploitation Windows et Linux.	
Talend	Un éditeur de solutions dédiées à la gestion des données d'entreprise depuis 2005 à Suresnes. La solution ETL de Talend est « Talend Open Studio for Data Intégration » qui est une solution Open source. Talend est une solution complète pour réaliser des projets ETL en temps record.	
Power BI	Une application développée par Microsoft qui permet de créer des tableaux de bord et des rapports. En plus de cela, Power BI offre également la possibilité de transformer les données avant de créer le tableau de bord.	
StarUML	Un outil de modélisation qui permet la conception de vues spécifiques d'un système en utilisant différents types de diagrammes. Cet outil de modélisation est particulièrement utile pour la conception de logiciels et d'applications informatiques.	

Tableau 11: Environnement logiciel

2.2 Langage de développement



Nom	Description	Logo
Flask	Un framework minimaliste, flexible et extensible qui permet de construire des applications web en utilisant le langage de programmation Python. Il permet aux développeurs de choisir les bibliothèques et les outils qui conviennent le mieux à leurs besoins, tout en offrant une structure de base pour construire des applications web.	
Angular	Un Framework client-serveur open source, développé par l'équipe du projet "Angular" chez Google, qui permet de créer des applications web modernes et dynamiques.	

Tableau 12: Langage de développement

3. Réalisation du tableau de bord

Nous sommes maintenant prêts à commencer la partie pratique de notre rapport, en nous concentrant sur la création d'un tableau de bord efficace. Un tableau de bord est un outil de gestion d'entreprise qui utilise divers indicateurs de performance pour aider les décideurs à anticiper les tendances et à prendre des décisions éclairées. Pour remplir cette fonction, un tableau de bord doit être bien structuré, organisé et conçu avec soin afin de maximiser son potentiel d'aide à la décision.

3.1 Connexion

Avant de passer à la création effective du tableau de bord, nous devons établir une connexion entre l'outil Power BI et la base de données Oracle utilisée par la société WorldSoft. Pour ce faire, nous allons suivre les étapes suivantes :

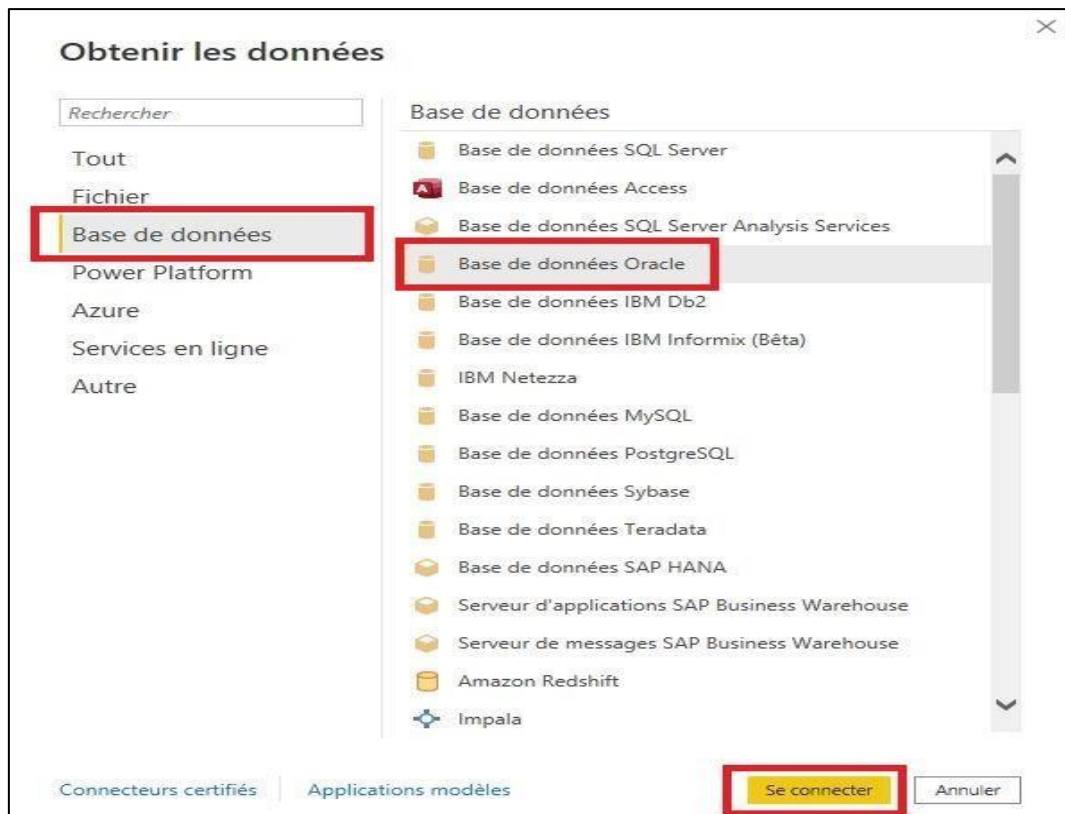


Figure 27: Interface 1 connexion Power BI à Oracle

Une fois que nous avons cliqué sur le bouton "Se connecter" de Power BI, l'interface suivante apparaît, nous permettant de saisir l'adresse du serveur de la base de données ainsi que d'autres informations nécessaires.



Figure 28: Interface 2 de connexion Power BI à Oracle

Une fois que nous avons établi la connexion à la base de données à l'aide de Power BI, nous allons récupérer les tables de dimensions et la table de faits nécessaires à la création de notre tableau de bord. La figure ci-dessous illustre cette étape de récupération.

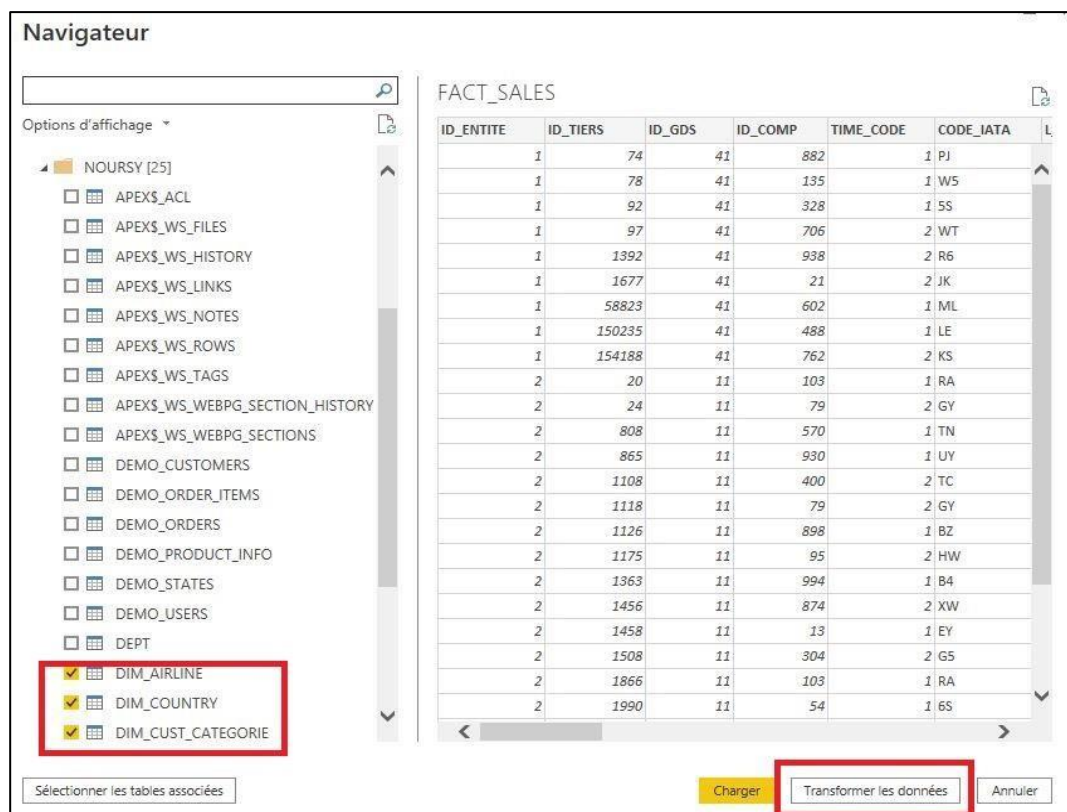


Figure 29: Récupération des tables sur Power BI

3.2 Visualisation du tableau de bord

La visualisation de données est un processus qui nous permet de convertir des données brutes en informations commerciales claires, attractives et utiles.

Pour cette raison, notre section visuelle est divisée en plusieurs rubriques :

- Accueil qui présente la situation générale des données à savoir nombre des agences, nombre de pays, nombre des compagnies aériennes ,nombre des GDS total
- Générale (vue d'ensemble globale)
- Agences
- Airlines
- Systèmes Global Distribution
- Pays

3.2.1 Ecran d'accueil

Lorsque l'utilisateur ouvre l'application, il est immédiatement présenté avec une vue d'ensemble des données clés, telles que le nombre d'agences, le nombre de pays, le nombre de compagnies aériennes et

le nombre total de GDS. Le design de l'interface est conçu de manière à refléter le domaine d'activité de l'entreprise, offrant une expérience utilisateur intuitive et cohérente.

La figure suivante nous montre l'écran d'accueil de notre application.

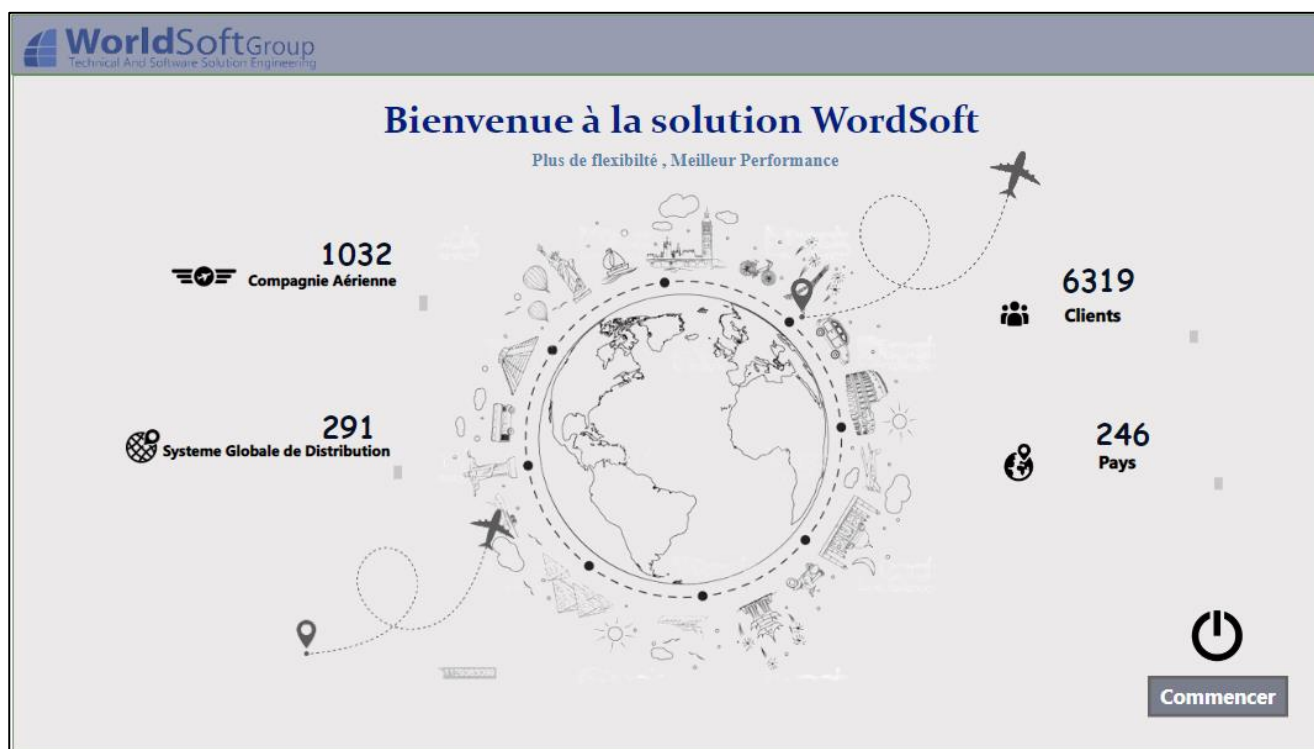


Figure 30: Ecran d'accueil de l'application

Une fois que le décideur a consulté la page d'accueil, il peut naviguer entre les différentes feuilles pour accéder à d'autres pages d'informations pertinentes.

3.2.2 Tableaux de bord de suivi général

Le tableau de bord suivant fournit des informations générales sur les ventes et les bénéfices par agence ou par pays, sous forme d'éléments graphiques variés. Ces informations indicatives sont présentées dans différents graphiques et tableaux pour offrir une vue d'ensemble complète.

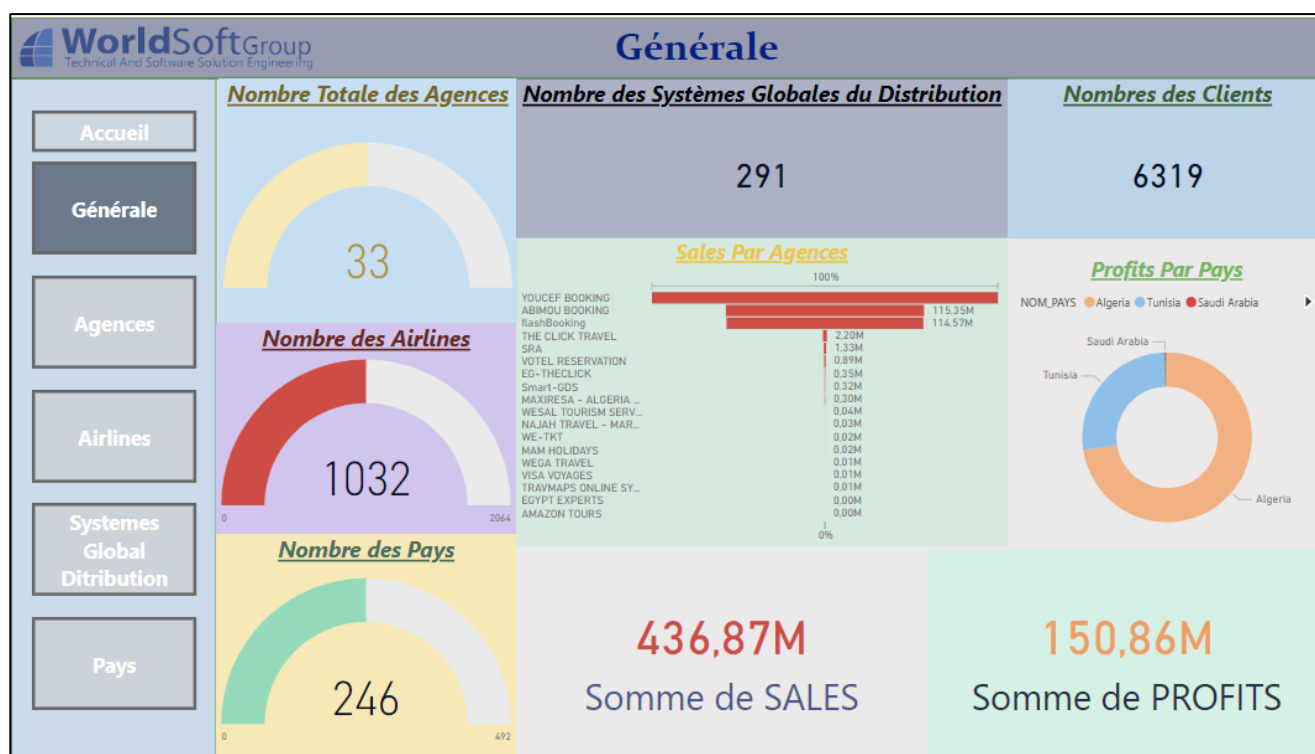


Figure 31: Rubrique vue d'ensemble

3.2.3 Tableaux de bord des « Agences »

Notre tableau de bord se concentre sur les informations liées aux agences (clients), en mettant en avant leur nationalité, leur nombre total, les bénéfices, la catégorie de l'agence, le nombre de GDS utilisés par chaque agence, ainsi que les compagnies aériennes associées à chaque agence. Les graphiques et tableaux présentés offrent une vue d'ensemble détaillée de ces informations clés.

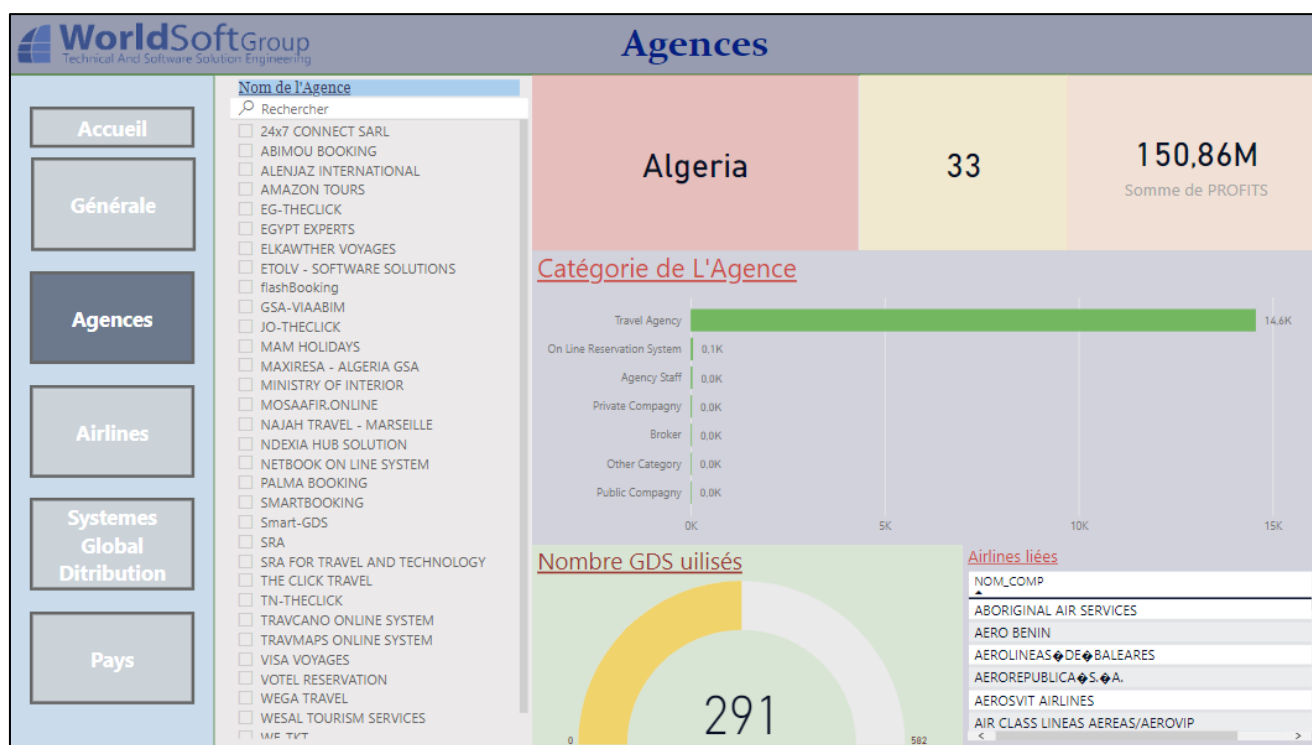


Figure 32: Rubrique des Agences

3.2.4 Tableaux de bord des « Airlines »

Notre tableau de bord se concentre également sur les informations relatives aux compagnies aériennes, en présentant tout d'abord une liste des différentes compagnies aériennes. Nous présentons ensuite des graphiques détaillant les informations clés associées à chaque compagnie aérienne, telles que le code IATA attribué, la nationalité, le nombre total de compagnies aériennes, ainsi que le type d'entreprise (domicile ou non).

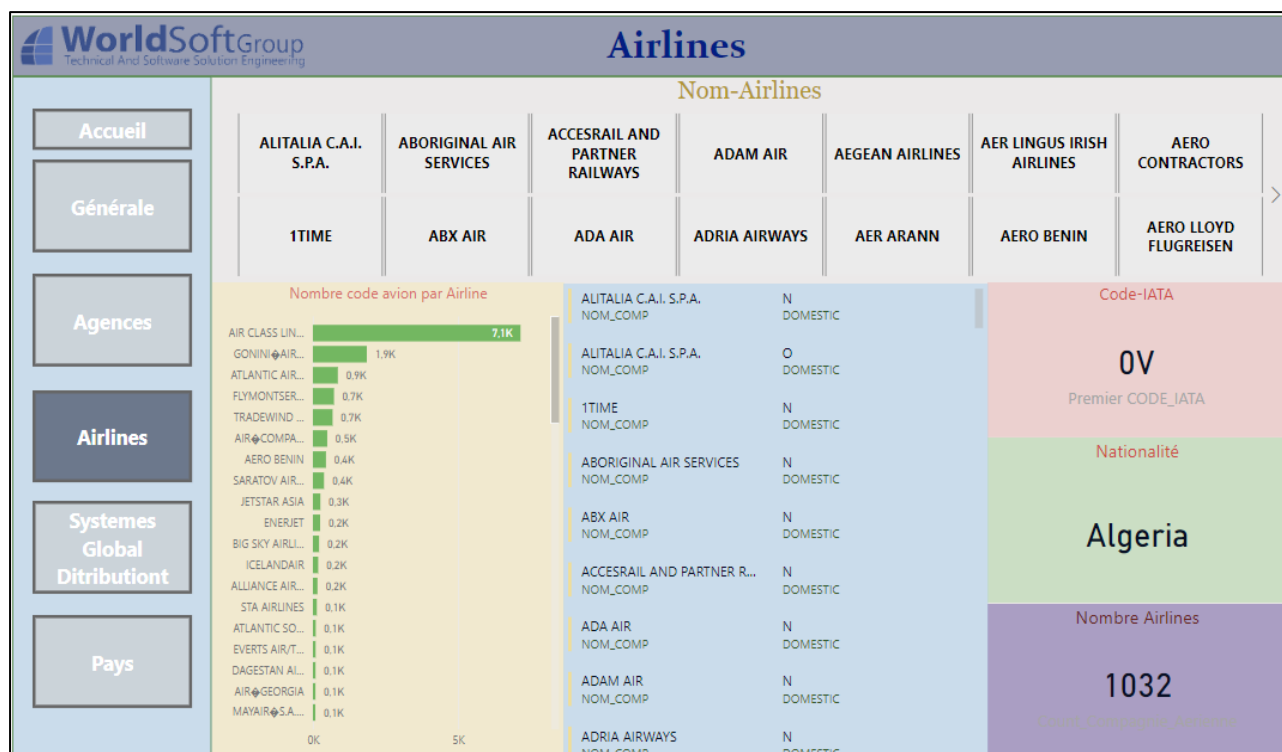


Figure 33: Rubrique Airlines

3.2.5 Tableaux de bord des « Systèmes Global Distribution »

Notre tableau de bord comporte également une liste des systèmes globaux de distribution de réservations (GDS), avec une présentation détaillée des informations clés associées à chaque GDS, telles que les bénéfices, le nombre total, le domaine de travail, la localisation avec une carte géographique, ainsi que les agences associées à chaque GDS.

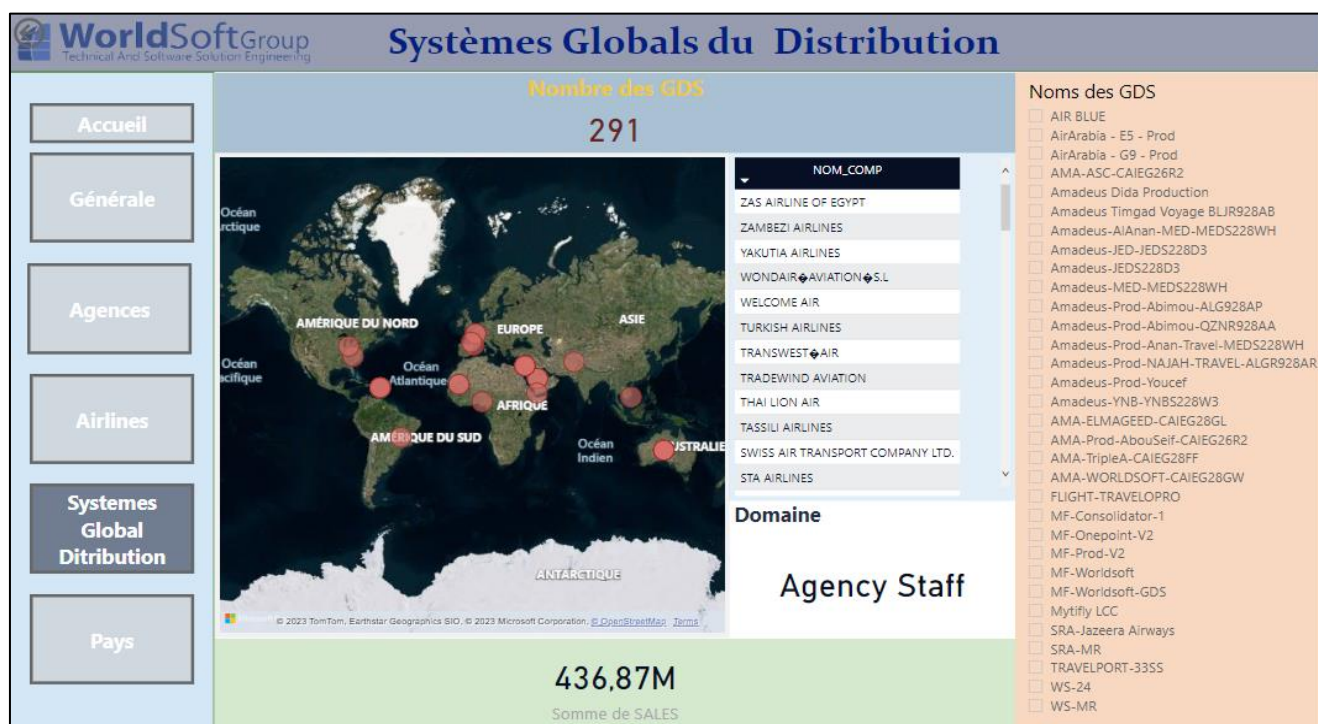


Figure 34: Rubrique des GDS

3.2.6 Tableaux de bord des « Pays »

Le tableau de bord que nous avons créé se concentre sur les aspects liés aux différents pays, en évaluant les profits générés, la moyenne des profits, le nombre d'agences présentes dans chaque pays, la devise utilisée, la liste des agences dans chaque pays et une carte graphique qui montre la localisation de chaque agence dans chaque pays. Nous avons également dressé une liste des pays inclus dans notre analyse.

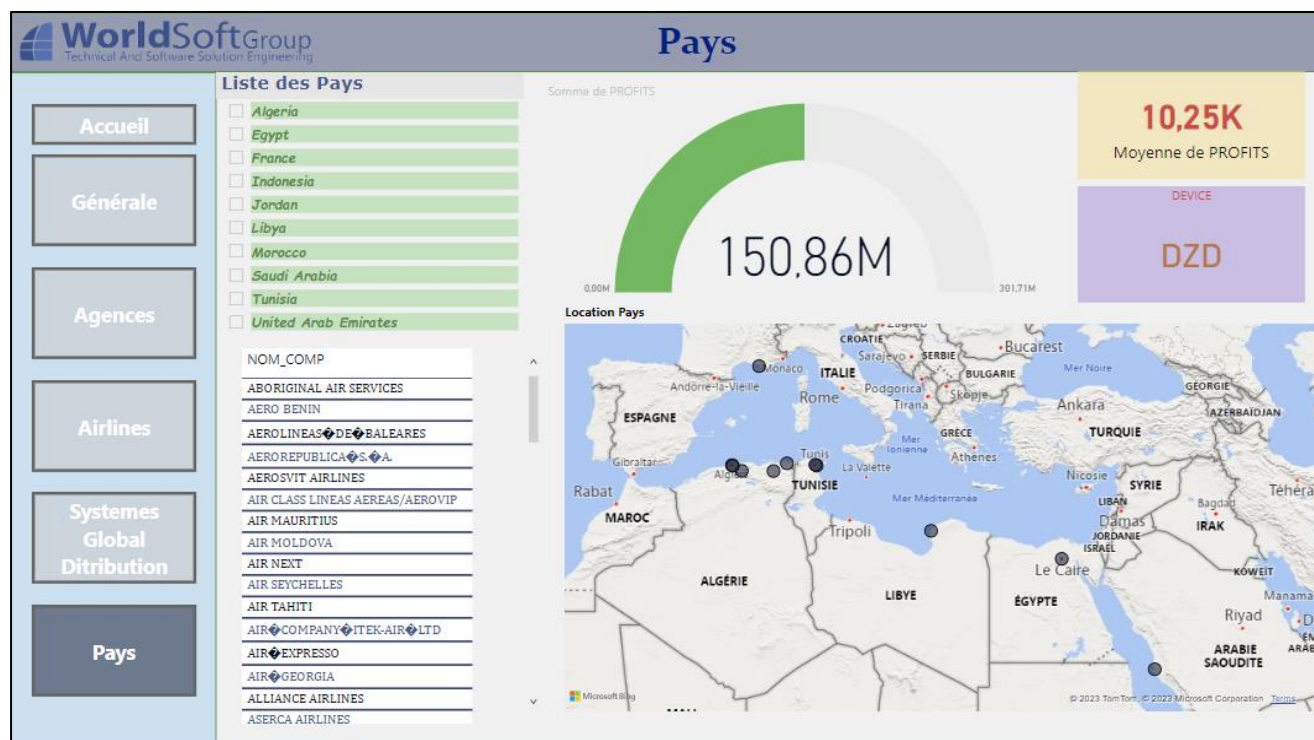


Figure 35: Rubrique des Pays

3.2 Déploiement sur Web

Une fois les tableaux de bord et le rapport générés, nous les intégrons dans un portail web dédié en utilisant l'élément iframe proposé par le service Power BI. Cet élément permet d'ajouter l'URL de l'objet à la page web, ce qui facilite la visualisation des données pour le décideur et améliore l'efficacité globale de l'application.

- **Authentification** : l'authentification de chaque utilisateur est requise pour accéder aux tableaux de bord et aux rapports. La figure suivante illustre la page d'authentification.

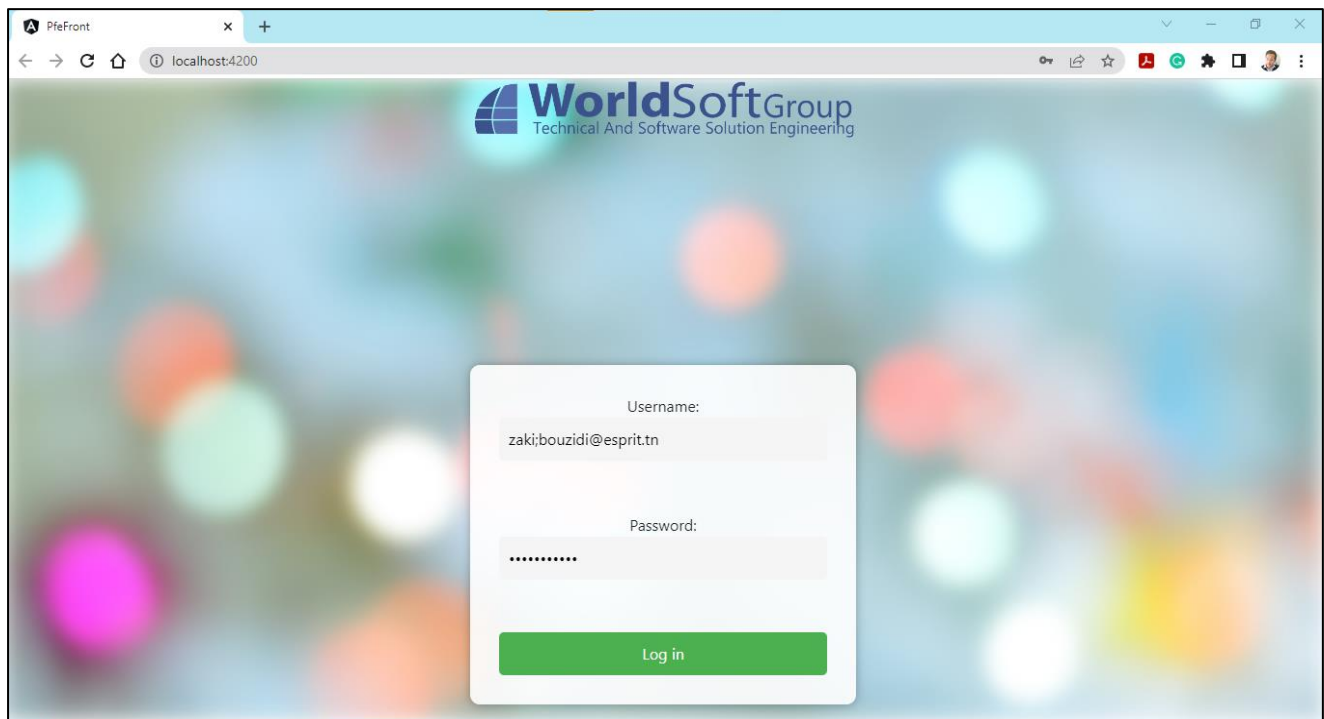


Figure 36: Authentification

- **Dashboards :** Une fois que l'on a été identifié, il est aisé d'accéder à tous les tableaux de bord de l'application BI spécialisée.



Figure 37: Application BI sur Web

Conclusion

Ce chapitre nous a permis de présenter les tableaux de bord, le rapport de relevé et leur incorporation dans l'application de manière très simple en utilisant une technologie innovante pour offrir au client une plateforme pratique et personnalisée.

Nous avons clarifié le travail final réalisé au cours de ce projet qui permet au client d'avoir une visibilité claire .

CONCLUSION GENERALE

Les entreprises modernes sont confrontées au défi de tirer parti des données pour démontrer leur valeur ajoutée. WorldSoft Group, une entreprise privée du secteur touristique en Tunisie et à l'étranger, a décidé d'intégrer un outil de Business Intelligence pour aider son manager à suivre l'évolution de ses activités financières.

Dans ce but, la société a demandé la conception d'un système décisionnel incluant un entrepôt de données ou Data Warehouse rassemblant toutes les données financières et commerciales de l'entreprise. Pour cela, des bases de données décisionnelles ont été utilisées pour concevoir un tableau de bord affichant les informations de vente de l'entreprise par zone, par client et par devise.

Le travail a été divisé en trois parties : une étude approfondie du métier de l'entreprise, une analyse détaillée de la source de données Oracle et la modélisation de l'entrepôt de données en utilisant le modèle en étoile.

Une fois le modèle créé, les données ont été alimentées et le cube OLAP a été réalisé pour préparer le tableau de bord contenant les indicateurs de vente. Enfin, le tableau de bord a été intégré sur la plateforme de la société à l'aide d'outils de développement web.

Ce projet a permis d'acquérir de multiples connaissances utiles et du savoir-faire en utilisant des outils open source et en se projetant dans le monde réel des entreprises.

Webographie

[1] https://www.oracle.com/webfolder/technetwork/tutorials/obe/db/12c/r1/olap/olap_cube/buildicubes.html

[2] <https://www.oracle.com/index.html>

[4] https://docs.oracle.com/cloud/help/fr/tax-reporting-cloud/AGTRC/admin_dimensions_period_100x5971b9f0.htm#AGTRC-GUID-F68A7A63-5F664E4A-9FF6-B84A9690DBCF

[5] <https://docs.oracle.com/en/database/data-integration/index.html>